OIPE
AUG 28 2006
PATENT & TRADEMARK OFFICE
IAP82

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of:	
Joseph McCrossan, et al.	Group Art Unit: 2615
Serial No.: 10/563,263	
Filed: June 6, 2006	
For: RECORDING MEDIUM, RECORDING METHOD, REPRODUCTION APPARATUS AND METHOD AND COMPUTER-READABLE PROGRAM	August 23, 2006 Costa Mesa, California 92626

PETITION TO MAKE SPECIAL

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Dear Sirs:

This Petition to Make Special is being submitted in accordance with 37 CFR §1.102(d) in order to accelerate examination of the above-identified application. Submitted below are items (A) through (D) as required pursuant to MPEP §708.02(VIII):

A. FEE

Submitted with this Petition to Make Special is the fee set forth in 37 CFR §1.17(h).

B. SINGLE INVENTION

In the event that the Office determines that all the claims presented are not obviously directed to a single invention, it is hereby submitted that the Applicants will make an election without traverse as a prerequisite to the grant of special status.

08/29/2006 GWORDOF1 00000091 192814 10563263
01 FC:1464 130.00 DA

C. PRE-EXAMINATION SEARCH

Submitted herewith is an International Search Report issued by the Japanese Patent Office in a corresponding foreign application having claims of the same scope to the claims currently pending in this application. (An English copy of the search report is attached.)

D. COPY OF REFERENCES

Submitted herewith is a copy of the following references which are cited in the foreign search report and which are deemed most closely related to the subject matter encompassed by the claims:

Document 1: WO 01/22729 A
Document 2: EP 0924 934 A
Document 3: US 6,104,706 A
Document 4: US 2003/117529 A1
Document 5: EP 1 035 735 A

E. DETAILED DISCUSSION

Provided next is a detailed discussion which points out, with the particularity required by 37 CFR §1.111(b) and (c), how the claimed subject matter is patentable over the cited references.

In the above documents, Document 1 was cited as of particular relevance to certain claims, particularly independent Claims 6, 13 and 14.

The present invention relates to the field of graphic display technology that reproduces a digital stream generated by multiplexing a video stream and a graphic stream. For example, as recording medium have increased their capacity to store data, it is possible to provide one or more movies in a single recording medium and to provide subtitles and other forms of graphic material that can be multiplexed with the video stream that can represent the moving picture. The graphic stream can include a plurality of display sets that are each made up of display control information in the graphics data. The present invention provides features to assist in the

reproduction of the content with a graphic display based on each of a plurality of display sets contained in the graphic stream.

Thus, referring to Claim 1, the graphic stream in the multiplexed video and graphics digital stream includes a plurality of display sets, each of which is used for graphics display. The display sets include a control segment and graphic data that is assigned an identifier. When an active period of the control segment in a specific display set overlaps with an active period of a control segment in an immediately preceding display set on a reproduction time axis, it is possible to distinguish by an identifier a particular control segment. The control segment can also be provided at the beginning of a display set with time information showing the decoding start time and time information showing a display start time.

Accordingly, the reproduction apparatus, in decoding of the video stream and the graphic stream, can identify and appropriately store decoded graphic data of a display set into a different area of an object buffer to facilitate timing and display of the information.

Each display set includes a control segment and graphics data that is assigned a unique identifier. The control segment includes time information which designates an active period of the control segment in the display set, on the reproduction time axis of the video stream. Since the active period is designated by the time information, if the active period of the control segment belonging to one display set and the active period of the control segment belonging to its succeeding display set are overlapped, then the graphics display composition processing is pipelined.

Thus, the background of the present invention is a data structure in which the active period of the control segment is designated by the time information and the graphics display is

realized based on this active period. In such a background, the present invention is characterized in the claims so that the identifier is assigned to the graphics data in the following manner:

> "if an active period of the control segment in the display set overlaps with an active period of a control segment in an immediately preceding display set on a reproduction time axis of the video stream, the identifier assigned to the graphics data in the display set differs from an identifier assigned to graphics data which is referenced by the control segment in the immediately preceding display set."

The present invention presents a distinctive feature that different identifiers are assigned to graphics data subjected to pipeline processing in order to remove problems caused by pipelining. The problems caused by pipelining are explained later.

The following discussion explains in detail the reason why the above claimed feature is neither anticipated by nor rendered obvious from the references.

D1 (WO 01/22729 A) discloses a closed caption tagging system for inserting tags into an audio or video television broadcast stream prior to or at the time of transmission. The tags contain command and control information. The receiver receives the broadcast stream and stores it on a storage device, and detects and processes the tags within the stored broadcast stream. Program material from the broadcast stream is played back to the viewer from the storage device. The receiver performs the appropriate actions in response to the tags. In this processing, three components that are a source, a transform, and a sink are pipelined (Figs. 8 and 9).

The source is a component that accepts data from encoders in a digital satellite receiver.

The transform is a component that executes a spatial transform (an image convolution or compression/decompression by data) and a temporal transform for the received data.

The sink is a component that consumes buffers, taking data from the transform, sending the data to the decoder, and then releasing the buffer for reuse.

The source, the transform, and the sink can be multithreaded and executed in parallel. Flow control of these components is exercised by the TmkPipeline class.

This reference D1 has some kind of data being processed by a pipeline. However, the target of the pipeline in this reference is the processes for the tags such as the source, the transform, and the sink, and not the realization of the graphics display corresponding to each display set of our present claims. Even if the tag processes in this reference are regarded as graphics display, this reference does not track any concept of removing problems caused by pipelining by means of assignment of identifiers to graphics data. Accordingly, this reference cannot be seed as a prior art reference that anticipates the present invention.

Also, reference D1 does not render obvious the present invention, because this reference merely discloses the idea of pipelining the tag processes in the receiver, and does not show the concept on which the above features are based, i.e. the concept of preventing, if an active period of the control segment in the display set overlaps with an active period of a control segment in an immediately preceding display set on a reproduction time axis of the video stream, the graphics data referenced by the control segment in the immediately preceding display set from being referenced by the control segment in the display set. Hence the present invention is unobvious from this reference.

D2 (EP 0924 934 A) discloses a technique in which each of a coding circuit for audio signals, a coding circuit for video signals, and a coding circuit for scene data output time information representing a decoding timing, a composition circuit outputs time information representing a composition timing, and a multiplexing circuit multiplexes the compressed data based on these time information. The video stream obtained by the multiplexing circuit includes

the reference clock value, and the time stamps and compressed data for audio, video, and scene data respectively (Fig. 26).

This reference D2 shows display timing of some kind of data designated by time information. However, this reference does not show any concept of removing problems caused by pipelining by a solution of assignment of identifiers to graphics data.

Also, this reference merely shows a concept of specifying the display timings of audio, video, and scene data using time stamps, and does not show the claimed features of preventing, if an active period of the control segment in the display set overlaps with an active period of a control segment in an immediately preceding display set on a reproduction time axis of the video stream, the graphics data referenced by the control segment in the immediately preceding display set from being referenced by the control segment in the display set. Hence the present invention is unobvious from this reference.

D3 (US 6 104 706 A) discloses a technique in which the receiver receives video data, audio data, and text/graphics from the sender and plays back them in real time. The real time playback is maintained by delaying the extraction of audio data accumulated on a FIFO buffer according to the average buffer delay time in the system.

This reference is similar to the present invention in that the display timing of some kind of data is adjusted. However, this reference does not show any concept of removing the problems caused by pipelining by means of assignment of identifiers to graphics data. Accordingly, this reference cannot be served as a prior art reference that anticipates the present invention.

Also, this reference cannot be served as a prior art reference that renders obvious the present invention. This is because this reference merely discloses the idea of delaying the

extraction of audio data accumulated on the FIFO buffer according to the average buffer delay time in the system, and does not show the concept on which the above features are based, i.e. the concept of preventing, if an active period of the control segment in the display set overlaps with an active period of a control segment in an immediately preceding display set on a reproduction time axis of the video stream, the graphics data referenced by the control segment in the immediately preceding display set from being referenced by the control segment in the display set. Hence the present invention is unobvious from this reference.

D4 (US 2003/117529 Al) discloses a technique of transmitting additional information such as graphics information or subtitles in addition to video information. This additional information is transmitted separately, so that the user may choose whether the additional information is to be displayed or not. Here, the video signal includes information relating to the duration for which the additional information is to remain on the display. According to this information, the additional information can be displayed exactly for the duration of the desired time. This is advantageous in trick modes such as Fast Forward.

Thus, D4 reference relates to a graphic display time being adjusted by using some kind of time information. However, this reference does not show any concept of removing the problems caused by pipelining by means of assignment of identifiers to graphics data. Accordingly, this reference cannot be served as a prior art reference that anticipates the present invention.

Also, this reference cannot render obvious the present invention, because the time information in this reference merely designates the duration for which the additional information such a graphic is to remain on the display, and does not show the concept on which the above claimed features are based, i.e. the concept of preventing, if an active period of the control segment in the display set overlaps with an active period of a control segment in an immediately

preceding display set on a reproduction time axis of the video stream, the graphics data referenced by the control segment in the immediately preceding display set from being referenced by the control segment in the display set. Hence the present invention is unobvious from this reference.

D5 (EP 1 035 735 A) discloses a packetization device for achieving packetization of a moving image code string by adding a RTP (Real Time Protocol) header to each unit composed of one or more video packets out of the video packets included in the moving image code string.

The D5 reference has time information added to a stream and recorded on a recording medium. However, this reference fails to provide any description about a graphic stream nor any concept of removing the problems, caused by pipelining, by means of assignment of identifiers to graphic data. Accordingly, this reference cannot be served as a prior art reference that anticipates the present invention.

Also, this reference cannot be served as a prior art reference that renders obvious the present invention. This is because the RTP header in this reference merely designates a time for real time processing, and does not show the concept on which the above claim features are based, i.e. the concept of preventing, if an active period of the control segment in the display set overlaps with an active period of a control segment in an immediately preceding display set on a reproduction time axis of the video stream, the graphics data referenced by the control segment in the immediately preceding display set from being referenced by the control segment in the display set. Hence the present invention is unobvious from this reference.

The above differences of the present invention as defined in the claims contribute to the following effects.

With the provision of the above features, different identifiers are assigned to graphic data in one display set and graphic data in its immediately preceding display set. Accordingly, even when the active periods of the control segments in the two display sets overlap with each other, the graphic data in the immediately preceding display set will not be overwritten by the graphic data in the display set. Therefore, the problems caused by such overwriting; i.e. the graphic which should be displayed later is displayed in place of the graphic which should be displayed earlier, will not occur. By assigning the identifiers in this way, the original display order of the graphics can be maintained. Hence a reproduction apparatus capable of processing display sets in parallel can make full use of its capability.

Also, two or more display sets can be processed in a pipeline even with a single processor for decoding graphics. Such pipeline processing increases decoding efficiency, without complicating the internal construction of the reproduction apparatus, thereby permitting a cost effective commercial product.

Claims 2-5 are dependent from Claim 1, and so are neither anticipated by nor rendered obvious from the references. Also, Claim 12 (recording method) is neither anticipated by nor rendered obvious from the references, for the same reason as Claim 1.

Claims 6-11 (reproduction apparatus), Claim 13 (program), and Claim 14 (reproduction method) have been amended. Claims 6, 13, and 14 to have been amended to include the features of Claim 1. As a result of this amendment, these claims are neither anticipated by nor rendered obvious from D1 for the same reason as Claim 1.

It is believed that applicant has satisfied the requirements for the request for Petition to Make Special and if there are any questions with regards to this matter, the undersigned attorney would appreciate a telephone conference and can be reached at the phone number listed below.

I hereby certify that this correspondence is being deposited with the United States Postal Service as First Class Mail in an envelope addressed to the Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450 on August 23, 2006.

By: Sharon Farnus

Sharon Farnus

Signature

Dated: August 23, 2006

Very truly yours,

SNELL & WILMER L.L.P.

Joseph W. Price
Registration No. 25,124
600 Anton Boulevard
Suite 1400
Costa Mesa, CA 92626
Telephone: (714) 427-7420
Facsimile: (714) 427-7799

PATENT COOPERATION TREATY

PCT

INTERNATIONAL SEARCH REPORT

(PCT Article 18 and Rules 43 and 44)

Applicant's or agent's file reference	FOR FURTHER ACTION	see Form PCT/ISA/220 as well as, where applicable, item 5 below.
P035118-P0		
International application No.	International filing date *(day/month/year)*	(Earliest) Priority Date *(day/month/year)*
PCT/JP2004/010153	09/07/2004	11/07/2003
Applicant		
MATSUSHITA ELECTRIC INDUSTRIAL CO., LTD.		

This International Search Report has been prepared by this International Searching Authority and is transmitted to the applicant according to Article 18. A copy is being transmitted to the International Bureau.

This International Search Report consists of a total of 4 sheets.

[X] It is also accompanied by a copy of each prior art document cited in this report.

1. **Basis of the report**
   a. With regard to the **language,** the international search was carried out on the basis of the international application in the language in which it was filed, unless otherwise indicated under this item.
      [ ] The international search was carried out on the basis of a translation of the international application furnished to this Authority (Rule 23.1(b)).
   b. [ ] With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, see Box No. I.

2. [ ] **Certain claims were found unsearchable** (See Box II).

3. [ ] **Unity of invention is lacking** (see Box III).

4. With regard to the **title,**
   [X] the text is approved as submitted by the applicant.
   [ ] the text has been established by this Authority to read as follows:

5. With regard to the **abstract,**
   [X] the text is approved as submitted by the applicant.
   [ ] the text has been established, according to Rule 38.2(b), by this Authority as it appears in Box No. IV. The applicant may, within one month from the date of mailing of this international search report, submit comments to this Authority.

6. With regards to the **drawings,**
   a. the figure of the **drawings** to be published with the abstract is Figure No. 25A
      [ ] as suggested by the applicant.
      [ ] as selected by this Authority, because the applicant failed to suggest a figure.
      [X] as selected by this Authority, because this figure better characterizes the invention.
   b. [ ] none of the figures is to be published with the abstract.

INTERNATIONAL SEARCH REPORT

International Application No
PCT/JP2004/010153

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 H04N5/92

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 H04N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, WPI Data, PAJ

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 01/22729 A (TIVO INC) 29 March 2001 (2001-03-29) page 3, line 16 - line 22 page 7, line 6 - page 10, line 17 page 11, line 7 - page 15, line 21	6-10,13, 14
A	page 16, line 1 - page 19, line 2 page 19, line 30 - page 20, line 35 page 25, line 30 - page 28, line 7 page 30, line 9 - line 26 page 39, line 24 - page 40, line 37 figures 1-9,11,12,17,19,20	1-5,12

-/--

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
5 November 2004	16/11/2004

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Fragua, M

Form PCT/ISA/210 (second sheet) (January 2004)

INTERNATIONAL SEARCH REPORT

International Application No
PCT/JP2004/010153

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	EP 0 924 934 A (NIPPON ELECTRIC CO) 23 June 1999 (1999-06-23) paragraph '0004! - paragraph '0006! paragraph '0009! - paragraph '0010! paragraph '0025! - paragraph '0027! paragraph '0033! - paragraph '0034! paragraph '0039! - paragraph '0042! paragraph '0045! - paragraph '0051! paragraph '0063! - paragraph '0065! paragraph '0068! paragraph '0073! paragraph '0077! paragraph '0079! - paragraph '0080! paragraph '0084! - paragraph '0085! paragraph '0094! paragraph '0096! - paragraph '0100! paragraph '0116! - paragraph '0119! figures 1,3-5,7,18-21,28-37,53	1-14
A	US 6 104 706 A (RICHTER ET AL) 15 August 2000 (2000-08-15) column 1, line 62 - column 2, line 67 column 7, line 24 - line 58 column 12, line 55 - line 65 figures 6,7a,7b	6-11,13, 14
A	US 2003/117529 A1 (DE HAAN) 26 June 2003 (2003-06-26) paragraph '0002! - paragraph '0015! paragraph '0028! - paragraph '0031! paragraph '0033! - paragraph '0036! figures 1-8	1,6, 12-14
A	EP 1 035 735 A (TOKYO SHIBAURA ELECTRIC CO) 13 September 2000 (2000-09-13) paragraph '0012! - paragraph '0014! paragraph '0017! - paragraph '0021! paragraph '0025! - paragraph '0027! paragraph '0033! - paragraph '0045! figures 1,4,6,8	1,6, 12-14

1

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/JP2004/010153

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 0122729 A	29-03-2001	AU 7706500 A	24-04-2001
		CN 1451234 T	22-10-2003
		EP 1214842 A1	19-06-2002
		JP 2003521851 T	15-07-2003
		WO 0122729 A1	29-03-2001
EP 0924934 A	23-06-1999	JP 3407287 B2	19-05-2003
		JP 11187398 A	09-07-1999
		CA 2256230 A1	22-06-1999
		EP 0924934 A1	23-06-1999
		US 6584125 B1	24-06-2003
US 6104706 A	15-08-2000	US 5995491 A	30-11-1999
		US 5623490 A	22-04-1997
		US 6738357 B1	18-05-2004
		AT 237894 T	15-05-2003
		AU 7206994 A	03-01-1995
		CA 2173355 A1	22-12-1994
		DE 69432524 D1	22-05-2003
		DE 69432524 T2	01-04-2004
		EP 1343290 A2	10-09-2003
		EP 0739558 A1	30-10-1996
		NZ 268754 A	28-07-1998
		WO 9429979 A1	22-12-1994
US 2003117529 A1	26-06-2003	AT 213113 T	15-02-2002
		CA 2200335 A1	06-02-1997
		DE 69619091 D1	21-03-2002
		DE 69619091 T2	02-10-2002
		DK 787404 T3	13-05-2002
		EP 0787404 A2	06-08-1997
		ES 2172663 T3	01-10-2002
		WO 9704591 A2	06-02-1997
		JP 10507607 T	21-07-1998
		PT 787404 T	31-07-2002
EP 1035735 A	13-09-2000	EP 1035735 A2	13-09-2000
		JP 2001148853 A	29-05-2001

(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 1 035 735 A2

(12) EUROPEAN PATENT APPLICATION

(43) Date of publication:
13.09.2000 Bulletin 2000/37

(51) Int. Cl.[7]: H04N 7/24

(21) Application number: 00104524.4

(22) Date of filing: 10.03.2000

(84) Designated Contracting States:
AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU MC NL PT SE
Designated Extension States:
AL LT LV MK RO SI

(30) Priority: 12.03.1999 JP 6712099
06.09.1999 JP 25192999

(71) Applicant:
KABUSHIKI KAISHA TOSHIBA
Kawasaki-shi, Kanagawa-ken (JP)

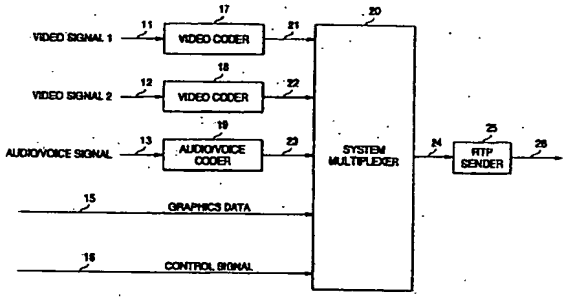
(72) Inventors:
- Kikuchi, Yoshihiro
  Minami-ku, Yokohama-shi, Kanagawa (JP)
- Masuda, Tadaaki
  Nerima-ku, Tokyo (JP)
- Nagai, Takeshi
  Saiwai-ku, Kawasaki-shi, Kanagawa-ken (JP)

(74) Representative: HOFFMANN - EITLE
Patent- und Rechtsanwälte
Arabellastrasse 4
81925 München (DE)

(54) Moving image coding and decoding apparatus optimised for the application of the Real Time Protocol (RTP)

(57) A moving image coding apparatus which has coders (17, 18, 19) for dividing an input moving image signal into a plurality of frames, dividing each of the frames into one or more image areas, compressing and coding the image areas, and outputting an area image code string, a system multiplexer (20) for separating frame header information indicating the coding mode, etc., of the frame frame from the frame frame and adding the frame header information to one or more coded area image code strings, and a sender (25) for collecting one or more area image code strings to which the frame header information is added, adding packet header information, putting into a packet, and sending the packet.

FIG. 1



EP 1 035 735 A2

BNSDOCID: <EP_____1035735A2_I_>

Description

DETAILED DESCRIPTION OF THE INVENTION

1. Field of the Invention

[0001] This invention relates to a moving image coding apparatus and a moving image decoding apparatus used with a system for compressing, coding, and multiplexing an image and voice and transmitting them via a network and particularly used with a system for transmitting a compressed image and voice on a packet-based network such as an intranet or the Internet.

2. Description of the Related Art

[0002] In video telephones, videoconference systems, digital television broadcasting, etc., a technique for compressing and coding a moving image and voice to less information amounts, multiplexing compressed moving image code string, voice code string, and data code string into one code string, and transmitting and storing the code string is used.

[0003] Techniques of motion compensation, discrete cosine transform (DCT), sub-band coding, pyramid coding, variable-length coding, etc., and systems provided by combining the techniques are developed. ISO MPEG1 and MPEG2 and ITU-T H.261, H.262, and H.263 exist as international standards for compressing and coding moving images, and ISO MPEG system, ITU-T H.221, H223, and the like exist as international standards for multiplexing code strings provided by compressing moving images and voice and audio signals and any other data. They are described in detail in document 1, "Multimedia coding no kokusaihyoujyun" edited and written by YASUDA Hiroshi, Maruzen (1994) and document 2, "MEPG-4 no subete" edited and written by MIKI, Kougyou chousakai (September 1998), and the like.

[0004] On the other hand, RTP (Realtime Transport Protocol) exists as a protocol for executing real-time transmission of a moving image code string provided by compressing and coding a moving image on a packet-based network such as an intranet or the Internet. The RTP is described in detail in document 3, Schulzrinne, Casner, Frederick, Jacobson RTP, "A Transport Protocol for Real Time Applications," RFC 1889, Internet Engineering Task Force (January 1996), and the like.

[0005] In addition to a fixed RTP header used in common, an RTP header proper to the compressing and coding technology can also be used as an RTP packet header. For example, the RTP headers for MPEG-1 and MPEG-2 are defined in document 4, D. Hoffman, G. Fernando, V. Goyal, M. Civanlar, "RTP Payload format for MPEG1/MEGP2 video," RFC 2250, Internet Engineering Task Force (January 1998).

[0006] Document 4 defines an RTP format for transmitting a previously multiplexed packet using an MPEG system and an RTP format proper to video/audio for entering a coded video/audio bit stream directly in an RTP packet.

[0007] In the former RTP format, one or more transport stream (TS) packets in an MPEG2 system in an RTP packet intact. Thus, if a transmission line error such as a packet loss occurs on a transmission line or medium for transmitting an RTP packet, it is made impossible to decode not only the lost RTP packet, but also the video bit stream in any other RTP packet to be decoded using the header information of the video bit stream contained in the lost RTP packet. Consequently, the transmission line error causes large degradation to occur in the decoded video signal; this is a problem.

[0008] On the other hand, as the latter RTP format, an RTP format extended for an MPEG video bit stream is used. FIG. 16 shows an example of the extended RTP format proper to MPEG video. In FIG. 16, f_[0,0], f_[0,1], f_[1,0], f_[1,1], DC, PS, T, P, C, Q, V, A, R, etc., is the same as information contained in a picture header in an MPEG video bit stream. Thus, the information contained in the picture header in the video bit stream is also entered in an RTP header of any other RTP packet than the RTP packet in which the picture header is entered, whereby if the RTP packet in which the picture header is entered is lost, in any other RTP packet, the information contained in the RTP header can be used for video decoding.

[0009] However, the extended RTP format involves the following problems:

(1) To prepare and transmit an RTP packet in a coding apparatus, processing of entering the header information contained in a video code string in an RTP packet header must be performed. After the RTP packet is received in a decoding apparatus, the information contained in the RTP header must be decoded and passed to a video decoding apparatus. The operation amounts increase because the steps are involved.

(2) The advantage of the extended RTP format can be provided on a network capable of transmitting RTP packets, such as an intranet or the Internet, but cannot be provided on a network incapable of transmitting RTP packets, such as a circuit switching network, since video code strings must be transmitted using any other multiplexing system other than the RTP.

[0010] As described above, to transmit packets undergoing system multiplexing in RTP packets in the coding apparatus for coding a moving image signal and transmitting the coded signal using an RTP packet, when the RTP packet containing important information such as the header information on a video bit stream is lost, this error also affects other RTP packets, causing large degradation to occur in the decoded moving

BNSDOCID: <EP_____1035735A2_I_>

image signal.

[0011] To use the RTP format proper to video coding, processing for entering the header information contained in a video code string in an RTP header becomes intricate. To connect a network capable of transmitting RTP packets also to a network incapable of transmitting RTP packets for transmitting a video code string, the advantage of the RTP extended header cannot be provided.

SUMMARY OF THE INVENTION

[0012] The invention has been made to solve the above problem, and therefore an object of the Invention is to provide a moving image coding apparatus and a moving image decoding apparatus for suppressing the adverse effect of an RTP packet loss when a moving image signal is coded and is transmitted using an RTP packet and simplifying processing of entering header information in an RTP header.

[0013] According to the invention, there is provided a moving image coding apparatus comprising coding means for dividing an input moving image signal into a plurality of screens (frames), dividing each of the screens (frames) into one or more image areas, compressing and coding the image areas, and outputting an area image code string, means for separating screen (frame) header information indicating the coding mode, etc., of the screen (frame) from the screen and adding the screen (frame) header information to one or more coded area image code strings, and conversion-to-packet means for collecting one or more area image code strings to which the screen header information is added, adding packet header information, putting into a packet, and sending the packet.

[0014] According to the invention, there is provided a moving image decoding apparatus comprising reception means for receiving a moving image code string put into a packet, separation means for separating one or more area image code strings contained in each packet of the moving image code string, area image decoding means for decoding the separated area image code string and outputting a decoded area image signal, screen decoding means for assembling the decoded area image signal for each screen (image frame) and outputting a decoded screen signal (decoded image frame signal), and means for generating a decoded moving image signal based on the decoded screen signal.

BRIEF DESCRIPTION OF THE DRAWINGS

[0015] In the accompanying drawings:

FIG. 1 is a block diagram of a coding apparatus according to a first embodiment of the invention;
FIG. 2 is a drawing to show the hierarchical structure of a video code string;
FIGS. 3A to 3D are drawings to describe video packets;
FIG. 4 is a block diagram to show the configuration of a system multiplexer;
FIG. 5 is a drawing to show the formats of an RTP packet header and payload;
FIGS. 6A to 6E are drawings to show the relationships among RTP packet, sync layer packet, and video bit stream;
FIG. 7 is a block diagram of a decoding apparatus corresponding to the coding apparatus in FIG. 1;
FIG. 8 is a block diagram to show the configuration of a system demultiplexer;
FIG. 9 is a block diagram of a coding apparatus according to a second embodiment of the invention;
FIG. 10 is a drawing to show the format of a video RTP packet;
FIGS. 11A to 11E are drawings to show the relationship between RTP packet and video bit stream;
FIG. 12 is a block diagram of a decoding apparatus corresponding to the coding apparatus in FIG. 9;
FIG. 13 is a block diagram of a coding apparatus according to a third embodiment of the invention;
FIG. 14 is a block diagram of a decoding apparatus corresponding to the coding apparatus in FIG. 13;
FIGS. 15A to 15E are drawings to show time stamp formats to describe a fourth embodiment of the invention;
FIG. 16 is a drawing to show an RTP format in a related art;
FIGS. 17A to 17C are drawings to show examples of RTP packet division prohibited according to RTP packet division rules;
FIG. 18 is a block diagram to show a coding apparatus for generating information and a medium for recording the information according to the invention;
FIG. 19 is a block diagram to show an information record medium and a decoding apparatus for decoding the information according to the invention;
FIG. 20 is a flowchart to show information recording and preparation processing according to the invention; and
FIG. 21 is a block diagram to show an example of a wireless moving image transmission system incorporating the coding apparatus and the decoding apparatus according to the invention.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

[0016] Referring now to the accompanying drawings, there are shown preferred embodiments of the invention.

(First embodiment)

[0017] FIG. 1 shows the configuration of a coding

BNSDOCID: <EP_____1035735A2_I_>

apparatus according to a first embodiment of the invention. Video signals 11 and 12 and an audio/voice signal 13 input from input means for inputting a moving image, such as a camera or a videocassette recorder (VCR), and converted into digital signals are input to video coders 17 and 18 and an audio/voice coder 19 respectively. Graphics data 15 and a control signal 16 for performing control are input to a system multiplexer 20.

[0018] The video signals 11 and 12 are compressed and coded by the first and second video coders 17 and 18 and are input to the system multiplexer 20 as first and second video code strings 21 and 22. The audio/voice signal 13 is compressed and coded by the audio/voice coder 19 and is input to the system multiplexer 20 as an audio/voice code string 23.

[0019] The video code strings 21 and 22, the audio/voice code string 23, the graphics data 15, and the control signal 16 are multiplexed by the system multiplexer 20 to generate a system code string 24. An RTP sender 25 puts the system code string 24 into an RTP packet and sends it as an RTP packet 26.

[0020] The video coders 17 and 18 performs highly efficient compression coding of a moving image signal by using DCT, quantization, variable-length coding, inverse quantization, inverse DCT, motion compensation, etc. That is, the moving image signal is divided into a plurality of frames, for example, frames and each frame is divided into one or more image areas, namely, blocks. The blocks are compressed and coded in accordance with a coding mode such as an intracoding mode or an intercoding mode to prepare a block coding string (image area coding string). Such processing is described in detail in document 2, etc., and therefore only the topics related to the invention will be discussed.

[0021] The number of video signals and that of video coders may be one or may be two or more as in the example in FIG. 1. To code a plurality of video signals, for example, before a moving image signal is coded, it can also be divided into a plurality of video objects such as a human figure and a background for inputting and coding the objects separately.

[0022] To handle such video objects, video bit stream has a hierarchical structure as shown in FIG. 2. The layer corresponding to the general sequence of a moving image is called VS (Visual Object Sequence) and one or more VOs (Visual Objects) exist in the VS. For example, if a human figure exists in a background, successive motion of only the human figure can be described as one VO, and a sequence of only the background can also be described individually. Further, each VO has a layer called VOL (Video Object Layer) under the VO. The VOL is a layer for giving a plurality of spatial resolutions or temporal resolutions to the VO; it is provided for performing spatio/temporal scalability coding. VOP (Video Object plane) at the lowest layer corresponds to a conventional frame and means data at "one instant" in each resolution of each VO (snap shot). A layer called GOV (Group of VOP) containing time information, etc., for executing random access exists between the VOL and VOP as an option.

[0023] If a code string is sent via a transmission line or medium where a bit error or a packet loss occurs, the following mechanism is adopted for video coding in order to reduce the adverse effect of the error:

[0024] As shown in FIG. 3A, the VOP is separated into units called video packets each consisting of several macro blocks (MBs). A marker for recovering synchronization (RM: Resynchronization marker) is added to the top of each video packet of a video code string, as shown in FIG. 3B.

[0025] FIGS. 3C and 3D are drawings to show header information of the video packet (VP header in FIG. 3B). The video packet header contains a flag called HEC (Header Extension Code). If the flag is "1," information of time code (MTB, VTI), VOP coding mode (VCP), intra DC VLC table change information (intra DC VLC threshold, IDVT), motion vector range information (VOP F code forward, VFF), etc., contained in the VOP header is also added to the video packet header, as shown in FIG. 3D.

[0026] FIG. 4 shows the configuration of the system multiplexer 20. The system multiplexer 20 is made up of access unit generators 31a to 31e and a sync layer packet (SL-PDU) generator 32. The access unit generators 31a to 31e separate input code strings 21, 22, 23, 15, and 16 into predetermined units called access units. For example, the video code string may be separated into access units in VOP units. The number, time stamp, and the like for identifying the code string are added to each access unit.

[0027] The access units are input to the sync layer packet generator 32, which then generates sync layer packets (also called SL-PDU) as a system code string 24. For the sync layer packets, the access units may be used intact or the access units may be divided into further fine units. The system code string 24 consisting of the generated sync layer packets is sent to the RTP sender 25 in FIG. 1, which then generates an RTP packet 26.

[0028] FIG. 5 shows an example of the generated RTP packet 26. It shows the RTP packet separated every 32 bits; 00 to 31 on the horizontal axis indicate bit positions of the RTP packet separated every 32 bits. In the figure, fields of V, P, X, ... CSRC shown as RTP Header provide the RTP header (RTP fixed header). This topic is described in detail in document 3 and therefore will not be discussed again in detail.

[0029] The sync layer packet generated by the sync layer packet generator 32 is entered in RTP payload in FIG. 5. In the RTP payload, first a sync layer packet header (SL-PDU header) is placed, followed by sync layer packet payload (SL-PDU payload), the contents of the sync layer packet. If the number of bits of the RTP payload is not a multiple of 32, a bit string called RTP padding may be added to the end of the RTP payload so that the number of bits of the RTP packet becomes a

BNSDOCID: <EP_____1035735A2_I_>

multiple of 32.

[0030] For some information in the RTP header, the information contained in the sync layer packet header may be used intact. For example, time stamp information in the sync layer packet header may be used as time stamp information in the RTP header. In this case, the time stamp may be removed from the sync layer packet header.

[0031] The access unit generators 31a to 31e and the sync layer packet generator 32 divide the video code string based on the following rules:

(1-1) Each header above the GOV in the hierarchical structure in FIG. 2 must be placed at the top of the sync layer packet payload (just after the sync layer packet header) or just after the higher-layer header;

(1-2) a higher-layer header than the header placed at the top of the sync layer packet payload must not exist at an intermediate point of the payload;

(1-3) if one or more heads exist in the sync layer packet payload, the payload must always begin with the header; and

(1-4) header must not be divided across sync layer packets.

[0032] FIGS. 6A to 6E are drawings to show examples of RTP packets generated as a result of generating sync layer packets based on the rules.

[0033] FIG. 6A shows the RTP packet in the beginning portion of a video bit stream sequence. According to rule (1-1), the VS (Visual Object Sequence) header, the VO (Visual Object) header, and the VOL (Video Object Layer) header above the GOV are successively placed just after the sync layer packet header. If the VS header, the VO header, or the VOL header, which has a small code amount, is divided across sync layer packets, RTP packets, code amount overhead caused by the RTP head or the sync layer packet header grows and the code amount increases. The header information pieces are entered in one RTP packet as shown in FIG. 6A, whereby the overhead caused by the RTP header or the sync layer packet header is reduced and an increase in the code amount is suppressed.

[0034] FIGS. 6B and 6C show examples of entering one video packet in one RTP packet. When the packet loss rate of the transmission line for sending a code string is high, if each video packet is entered in one sync layer packet, RTP packet, even if a packet loss occurs, only one video packet is lost, so that error resilience is improved. As previously described with reference to FIG. 3D, if video coding is performed so that a part of the VOP header information is entered in the video packet header, the information can be used to decode a moving image if the RTP packet containing the VOP header is lost. In the example, the access unit generators 31a to 31e may divide access units for each VOP and further the sync layer packet generator 32 may divide sync layer packets for each video packet.

[0035] FIG. 6D shows an example of entering a plurality of video packets in one RTP packet. If too fine division into RTP packet is executed, overhead caused by the RTP header or the sync layer packet header grows. Thus, if the bit rate of the transmission line is low, a plurality of video packets may be thus entered in

[0036] one RTP packet. FIG. 6E shows an example of entering a plurality of VOPs in one RTP packet. In doing so, the overhead caused by the RTP head, the SL-PDU header can be reduced more than that in FIG. 6D.

[0037] Padding bits may be added the end of each RTP packet in FIGS. 6A to 6E so that the RTP packet length becomes a multiple of 32 bits.

[0038] FIG. 7 is a block diagram to show the configuration of a decoding apparatus corresponding to the coding apparatus in FIG. 1. A code string 101 sent via a transmission line or a storage medium (not shown) is input to an RTP receiver 102. The RTP receiver 102 decodes the time stamp, the sequence number, etc., in the RTP packet header and outputs a sync layer packet 103 to a system demultiplexer 104.

[0039] If the RTP sender 25 removes some information of the time stamp, etc., in the sync layer packet header and enters the remaining information in the RTP header in, the RTP receiver 102 restores the removed sync layer packet header information to the original based on the decoded time stamp from the RTP header.

[0040] If a packet loss of RTP packet or reversal of the packet arrival order occurs on the transmission line, the received RTP packet sequence numbers do not become serial or are reversed, thus the packet loss, etc., can be detected. The RTP receiver 102 may restore the reversed RTP packet order to the correct order or feed back the detected packet loss rate, etc., to the coder as RTCP information (not shown).

[0041] FIG. 8 is a block diagram to show the configuration of the system demultiplexer 104. First, a sync layer packet decoder 105 decodes an access unit based on the sync layer packet header information in the input sync layer packet 103. If the sync layer packet generator 32 divides one access unit into a plurality of sync layer packets, a sync layer packet decoder 105 assembles the sync layer packets into one original access unit. The generated access units are classified according to the type (video, audio/voice, graphics, control signal) and are output to corresponding access unit decoders 106a to 106e. The access unit decoders 106a to 106e decode the access unit headers and output first and second video code strings 121 and 122, an audio/voice code string 123, graphics data 115, and a control signal 116.

[0042] First and second video decoders 117 and 118 and an audio/voice decoder 119 decode the video code strings 121 and 122 and the audio/voice code string 123 respectively and output first and second video reconstruction signals 131 and 132 and an

BNSDOCID: <EP_____1035735A2_I_>

audio/voice reconstruction signal 133 respectively as reconstruction signals.

[0043] If the RTP receiver 102 detects a packet loss of RTP packet, it may send a signal 107 indicating occurrence of a packet loss to the system demultiplexer 104. The system demultiplexer 104 may input the signal 107 to the sync layer packet decoder 105 and for the packet where the packet loss occurred, a signal indicating occurrence of the packet loss (not shown) may be sent to the access unit decoders 106a to 106e instead of sending the access unit. Each of the access unit decoders 106a to 106e may send a signal indicating occurrence of the packet loss (not shown) to the video decoder 117 or the audio/voice decoder 119 based on the signal 107.

[0044] The video decoder 117 may perform the following decoding processing based on the sent signal indicating occurrence of the packet loss: For example, assume that video code string is divided for each video packet and RTP packet is generated, as shown in FIGS. 6B and 6C. Also, assume that the video packet header of the video packet in FIG. 6C contains some information of the VOP header as previously described with reference to FIG. 3C. If occurrence of packet loss in the RTP packet containing the VOP header in FIG. 6B is detected, to decode the video packet in the RTP packet in FIG. 6C, the video packet is decoded based on the information of the VOP header contained in the video packet header in place of the VOP header information. In doing so, if the RTP packet containing the VOP header is lost, the video code string contained in any other RTP packet can be decoded correctly.

[0045] According to the embodiment, VOP header information is added in the corresponding video coder 17 or 18 or the audio/voice coder 19 to the VOP header in FIG. 3 and is multiplexed in the system multiplexer 20. The packet header information is added to image code string in the RTP sender 25.

(Second embodiment)

[0046] FIG. 9 shows the configuration of a coding apparatus according to a second embodiment of the invention. Parts identical with those previously described with reference to FIG. 1 are denoted by the same reference numerals in FIG. 9 and only the differences from the coding apparatus of the first embodiment will be discussed. The coding apparatus of the second embodiment differs from that of the first embodiment in that it does not include the system multiplexer in the first embodiment, that first and second code strings 21 and 22, an audio/voice code string 23, graphics data 15, and a control signal 16 are input to RTP senders 151, 152, 153, 154, 155, and 156, and that RTP packets 161, 162, 163, 164, 165, and 166 are also output separately. The RTP packets are multiplexed on an IP packet layer (not shown).

[0047] FIG. 10 shows an example of an RTP packet corresponding to a video code string. The RTP header fields are given the same names as the information pieces contained in the RTP header of the RTP packet in FIG. 5, but they differ partially in meaning.

[0048] A partial code string provided by dividing the video code string is entered in RTP payload in FIG. 10. The video code string is divided based on the following rules:

(2-1) Each header above the GOV in the hierarchical structure in FIG. 2 must be placed at the top of the RTP payload (just after the RTP header) or just after the higher-layer header;
(2-2) a higher-layer header than the header placed at the top of the RTP payload must not exist at an intermediate point of the payload;
(2-3) if one or more heads exist in the RTP payload, the payload must always begin with the header; and
(2-4) video header must not be divided across RTP packets.

[0049] FIGS. 11A to 11E are drawings to show examples of RTP packets generated by dividing a video bit stream based on the rules (2-1) to (2-4). FIG. 11A shows the RTP packet in the beginning portion of the video bit stream sequence. According to rule (2-1), the VS (Visual Object Sequence) header, the VO (Visual Object) header, and the VOL (Video Object Layer) header above the GOV are successively placed just after the RTP header.

[0050] If the VS header, the VO header, or the VOL header, which has a small code amount, is divided across RTP packets, code amount overhead caused by the RTP header grows and the code amount increases. Then, the header information pieces are entered in one RTP packet as shown in FIG. 11A, whereby the overhead caused by the RTP header is reduced and an increase in the code amount is suppressed.

[0051] FIGS. 11B and 11C show examples of entering one video packet in one RTP packet. When the packet loss rate of the transmission line for sending a code string is high, if each video packet is entered in one RTP packet, even if a packet loss occurs, only one video packet is lost, so that error resistance is improved. As previously described with reference to FIG. 3D, if video coding is performed so that a part of the VOP header information is entered in the video packet header, the information can be used to code a moving image if the RTP packet containing the VOP header is lost.

[0052] FIG. 11D shows an example of entering a plurality of video packets in one RTP packet. If too fine division into RTP packet is executed, overhead caused by the RTP header grows. Thus, if the bit rate of the transmission line is low, a plurality of video packets may be thus entered in one RTP packet.

[0053] FIG. 11E shows an example of entering a plurality of VOPs in one RTP packet. In doing so, the

BNSDOCID: <EP_____1035735A2_I_>

overhead caused by the RTP header can be reduced more than that in FIG. 11D.

[0054] Padding bits may be added the end of each RTP packet in FIGS. 11A to 11E so that the RTP packet length becomes a multiple of 32 bits. As the information pieces of the RTP header, the following may be used:

[0055] For the time stamp shown in FIG. 10, the time stamp contained in the video code string may be used intact or may be used with only the bit format changed. If the time stamp in the video code string is variable-length code, it may be converted into fixed-length code. If only one VOP header is contained in the video code string in the RTP packet as in FIG. 11A or 11C, the time stamp contained in the VOP header or the time stamp whose format is changed is used. If more than one VOP header is contained as in FIG. 11E, the time stamp of the first VOP header may be used. If no VOP header is contained as in FIG. 11C, the time stamp of the VOP header to which the video packet belongs is used.

[0056] The M bit in FIG. 10 may be set, for example, as follows:

(3-1) M is set to 1 only for the RTP packet containing a GOV header and the RTP packet containing a VOP header of VOP (I-VOP) undergoing intraframe coding; M is set to 0 for other RTP packets.
(3-2) M is set to 1 only for the last RTP packet if one VOP head is divided across RTP packets.
(3-3) M is set to 1 only if more than one VOP head is contained in an RTP packet.
(3-4) M is set to 1 only if more than one video packet is contained in an RTP packet.

[0057] FIG. 12 is a block diagram to show the configuration of a decoding apparatus corresponding to the coding apparatus in FIG. 9. Parts identical with those previously described with reference to FIG. 7 are denoted by the same reference numerals in FIG. 12 and only the differences from the decoding apparatus in FIG. 7 will be discussed. The decoding apparatus in FIG. 12 differs from that in FIG. 7 in that the RTP packets corresponding to video, audio/voice, graphics data, and control information are input to separate RTP receivers and are processed. The RTP packets are distributed to the corresponding RTP receivers based on port numbers, etc., on an IP layer (not shown).

[0058] If a packet loss of RTP packet or reversal of the packet arrival order occurs on the transmission line, the received RTP packet sequence numbers do not become serial or are reversed, thus the packet loss, etc., can be detected. The RTP receiver may restore the reversed RTP packet order to the correct order or feed back the detected packet loss rate, etc., to the coder as RTCP information (not shown).

[0059] If the RTP receiver 251, 252, or 253 detects an RTP packet loss, it may send a signal indicating occurrence of a packet loss (not shown) to the video decoder 117 or 118 or the audio/voice decoder 119.

[0060] The video decoder 117, 118 may perform the following decoding processing based on the sent signal indicating occurrence of the packet loss: For example, assume that video code string is divided for each video packet and RTP packet is generated, as shown in FIGS. 11B and 11C. Also, assume that the video packet header of the video packet in FIG. 11C contains some information of the VOP header as previously described with reference to FIG. 3C. If occurrence of packet loss in the RTP packet containing the VOP header in FIG. 11B is detected, to decode the video packet in the RTP packet in FIG. 11C, the video packet is decoded based on the information of the VOP header contained in the video packet header in place of the VOP header information. In doing so, if the RTP packet containing the VOP header is lost, the video code string contained in any other RTP packet can be decoded correctly.

[0061] According to the embodiment, VOP header information and packet header information added in the video coder 17 or 18 or the audio/voice coder 19 are added to image code string in the RTP sender.

(Third embodiment)

[0062] FIG. 13 shows the configuration of a coding apparatus according to a third embodiment of the invention. Parts identical with those previously described with reference to FIGS. 1 and 9 are denoted by the same reference numerals in FIG. 13 and only the differences will be discussed in detail.

[0063] First, control information 16 is input to a control information sender 1056. The control information 16 contains information indicating the coding system and mode applied when a video coder 17 compresses and codes a video signal 11, information indicating the coding system and mode applied when an audio/voice coder 19 compresses an audio/voice signal 13, and information indicating the RTP coding system and mode applied in RTP senders 151 and 153.

[0064] The information indicating the coding system and mode may include the following:

- o Video coding method (MPEG-1, MPEG-2, MPEG-4, H.261, H.263, JPEG, etc.,), profile level (main profile main level, simple profile level 1, etc.,), coding option mode type;
- o information indicating the number of pixels of one frame of video signal (CIF/QCIF/SIF/VGA, etc.,) and the numbers of horizontal and vertical pixels;
- o time resolution of video signal (Hz, etc.,);
- o coding bit rate;
- o coding delay;
- o RTP coding method and configuration, for example, meaning of RTP time stamp, resolution, meaning of marker bit, etc.,;
- o information as to which of video signal and

BNSDOCID: <EP_____1035735A2_I_>

audio/voice signal is not coded.

[0065] The input control information 16 is coded in the control information sender 1056 and is input to a decoding apparatus (described later) via a transmission medium (not shown) as a control information code string 1066. At the time, the decoding apparatus may always perform decoding based on the information indicating the coding method and mode sent with the control information code string 1066. Alternatively, the following negotiation operation may be performed via a transmission medium (not shown) between the coding apparatus and the decoding apparatus:

(1) If the sent the information indicating the coding method and mode indicates a coding method or mode that cannot be applied in the decoding apparatus, information indicating the fact is sent to the control information sender 1056. Then, the control information sender 1056 again sends a control information code string 1066 indicating a coding method and mode changed in the range in which the coding apparatus can adopt. Such operation is repeated until the coding method and mode that can be applied in the decoding apparatus are found.

(2) Pairs indicating candidates of coding methods and modes that can be adopted in the coding apparatus are built in the control information code string 1066 and the decoding apparatus selects a suitable coding method and mode and sends the information indicating the selected coding method and mode to the control information sender 1056.

[0066] The information indicating a coding method and mode contained in the control information 16 is also sent to the video coder 17, the audio/voice coder 19, and the RTP senders 151 and 153, and coding is performed based on the coding method and mode. If the negotiation operation is performed, the information indicating the coding method and mode determined by the negotiation operation is sent.

[0067] The video signal 11 and the audio/voice signal 13 are input to the video coder 17 and the audio/voice coder 19 respectively and video coding and audio/voice coding are performed based on the coding method and mode indicated on the information sent from the control information sender 1056, then a video code string 21 and an audio/voice code string 23 are output.

[0068] The operation of the video coder 17 and the audio/voice coder 19 is similar to that in the coding apparatus in the first and second embodiments. The structure of the video code string 21 is also similar to that in the first and second embodiments, as shown in FIG. 3.

[0069] The video code string 21 and the audio/voice code string 23 are input to the RTP senders 151 and 153, and RTP coding is performed based on the coding method and mode indicated on the information sent from the control information sender 1056.

[0070] The RTP sender 151 divides the video code string 21 into packets in accordance with one determined rule, adds RTP header information containing a time stamp, etc., and generates RTP packet, then outputs as an RTP code string 162. Although dividing the video code string 21 into packets and getting information of the time stamp, etc., for RTP header generation may be performed while the video code string 21 is being analyzed, packet length information and time stamp information (not shown) may be sent from the video coder 17 to the RTP sender 151 and dividing into packets and RTP header generation may be performed based on the information. This eliminates the need for the RTP sender 151 to analyze the video code string 21, so that processing is reduced.

[0071] FIG. 14 is a block diagram to show the configuration of the decoding apparatus corresponding to the coding apparatus in FIG. 13.

[0072] First, a control information code string 1166 received via a transmission line or a storage medium (not shown) is input to a control information receiver 1156 and control information 136 concerning the coding method and mode used in the coding apparatus is decoded and output. At the time, the negotiation operation may be performed between the decoding apparatus and the control information sender 1056 for determining the coding method and mode, as described in the operation description of the coding apparatus in FIG. 13. Of the decoded and determined control information, the information concerning the coding method and mode of the video signal and that concerning the coding method and mode of the audio/voice signal are input to a video decoder 117 and an audio/voice decoder 119 respectively. The information concerning the coding method and mode of the RTP code strings is input to RTP receivers 251 and 253.

[0073] The RTP code strings 251 and 253 received via a transmission line or a storage medium (not shown) are received at the RTP receivers 251 and 253, and RTP decoding is performed, then a video code string 121 and an audio/voice signal code string 123 are output. The operation of the RTP receiver 251 and that of the RTP receiver 253 correspond to the operation of the RTP sender 151 and that of the RTP sender 153 respectively.

[0074] The video code string 121 and the audio/voice signal code string 123 are input to the video decoder 117 and the audio/voice decoder 119 respectively, which then perform video decoding and audio/voice decoding and output a video reconstruction signal 131 and an audio/voice reconstruction signal 133. The decoding operation of the video decoder 117 and that of the audio/voice decoder 119 correspond to the coding operation of the video coder 17 and that of the audio/voice coder 19 in the coding apparatus previ-

BNSDOCID: <EP_____1035735A2_I_>

ously described with reference to FIG. 13. They are similar to those of the decoders in the decoding apparatus of the first and second embodiments and therefore will not be discussed again in detail.

[0075] In the third embodiment, graphics data can also be transmitted and a plurality of video signals can also be coded and transmitted as in the first and second embodiments. In this case, separate RTP senders code and transmit the graphics data and a plurality of video signals.

[0076] In the embodiment, the RTP senders code the video code string and the audio/voice code string separately, but as in the first embodiment, first, system multiplexer 20 may multiplex the video code string and the audio/voice code string, then RTP sender may perform RTP coding. In this case, the control information sender may code only control signal 16 or new control information may be provided aside from the control information 16 and may be coded by the control information sender.

[0077] Sync layer packet (SL-PDU) generator 32 in the multiplexer 20 may only divide code strings output from access unit generators 31a to 31e into smaller packets as required without adding any header information. In this case, the SL-PDU header in the RTP format in FIG. 5 does not exist and only SL-PDU payload to which RTP padding is added as required exists in RTP payload.

[0078] In the above-described embodiment, the sequence number and the time stamp in the RTP header may begin with a random number. If they are set to determined initial values, such as 0, the possibility that a third party may find the first RTP packet in a video audio sequence by finding the initial value and may decode RTP code sting is high. If random numbers are set as the initial values, such a possibility is lowered and security is improved. If time stamp information is provided, for example, by converting from time stamp information in video code string, the time stamp in the video code string to which a random number is added may be adopted as the time stamp in the RTP header.

(Fourth embodiment)

[0079] A fourth embodiment of the invention is the same as the second and third embodiments in the basic configurations of coding apparatus and decoding apparatus; they differ only in time stamp field added to an RTP header and therefore only the differences will be discussed in detail.

[0080] FIGS. 15A to 15E are drawings to show examples of formats of time stamp multiplexed to RTP header (time stamp field in FIG. 10). In the MPEG-4 standard (refer to document 4), a time stamp in the format of combining an MTB (module_time_base) field provided by coding the time difference in second units in variable length and VTI (VOP_time_increment) indicating the time with a finer precision than seconds is used as time stamp in video code string.

[0081] FIG. 15A shows an example of using a variable-length-coded time stamp of MPEG4 video intact in time stamp field in RTP header. In this case, the time stamp information of the video code string in MPEG4 is put in the intact format, thus processing is simplified in such a system configuration comprising an MPEG4 video coding section and an RTP packet conversion section separately.

[0082] FIG. 15B shows a time stamp example wherein the absolute time from one time is used as a time base in second units without using the MTB provided by coding the time difference in second units in variable length as it is, and the VTI indicating a finer precision than seconds is represented in a fixed length of a proper number of bits. In this example, second units are also multiplexed directly to the RTP header in the absolute time. To use the time stamp information in the RTP header, processing is facilitated, stronger resistance to a packet loss can be provided, and further to use a header compressing technique of IP, UDP, and RTP beads together, higher efficiency can be provided.

[0083] That is, in the example in FIG. 15A, the time difference in second units is coded in variable length and thus to use the time stamp information in an RTP layer, processing of once decoding the variable-length code becomes necessary, but the time stamp in the example in FIG. 15B can be used directly without requiring the processing.

[0084] In the example in FIG. 15A, the MTB has a value other than zero only when the time stamp changes in second units. If a packet loss occurs in the packet by chance, the receiving party cannot sense time stamp change in second units and after this, a time stamp discrepancy in second units occur between the transmitting party and the receiving party all the while. In contrast, in the example in FIG. 15B, the elapsed time since one time is also represented by an absolute value in second units, so that such a discrepancy does not occur.

[0085] To use RTP on an intranet or the Internet, a technique called header compression may be used to avoid overhead of IP/UDP/RTP headers. The header compression is described in detail, for example, in document 5, "Compressing IP/UDP/RTP headers for Low-Speed Links," RFC 2508, Internet Engineering Task Force (Feb. 1999). In the header compression technique, information in the header field having the same value as the header information in the immediately preceding packet or information in the header field having a constant difference value from the header information in the immediately preceding packet usually is not transmitted and only when exceptional behavior occurs, the information in the field is sent.

[0086] In the RTP header, the time stamp field is also a filed to which header compression is applied. It is expected that in consecutive RTP packets, the values increase constantly and the difference value therebe-

BNSDOCID: <EP_____1035735A2_I_>

tween becomes constant. However, if representation of an MPEG4 video code string as in FIG. 15A is directly put as the time stamp in the RTP header for putting MPEG4 video on an RTP packet, the differences do not become constant in simple time stamp field difference processing between the preceding packet and the current packet, and the requirement of the header compression technique cannot be satisfied. As a result, the possibility that efficiency will not become very good is high even if header compression is executed.

[0087] Then, if the format as shown in FIG. 15B is used as a time stamp, such a problem does not arise and high compression efficiency can also be provided if IP/UDP/RTP header compression is executed.

[0088] In the format in FIG. 15C, serial number information (frame No.) of image frame is added to the format in FIG. 15B, whereby how many image frames are discarded when packet discard occurs can be easily known in addition to the above-described features of the format in FIG. 15C.

[0089] FIGS. 15D and 15E show examples of using composition time calculated from VTI and MTB. The composition time is provided by adding VTI representing the time with a finer precision than seconds to accumulation of the differences in second units represented by MTB. In the examples, the time stamp field in the RTP header can be represented flat without providing a more finely divided structure, so that RTP header processing is facilitated. In this case, the features that if header compression is executed, high compression efficiency can be provided and that if a packet loss occurs, the time stamp discrepancy between the transmitting and receiving parties does not occur as in the formats in FIGS. 15B and 15C are not impaired.

[0090] The formats in FIGS. 15D and 15E differ in representation precision of the composition time. In the format in FIG. 15D, the composition time is represented with a predetermined precision and in the format in FIG. 15E, the composition time is represented with the same precision as the representation precision of VTI in the video code string. In the format in FIG. 15D, for example, the representation precision may be made the same as the system clock precision of the coding apparatus and the decoding apparatus or may be made the same as the precision of the clock used on the network. In the example in FIG. 15E, the information indicating the representation precision may be contained in the control information and is sent from the coding apparatus to the decoding apparatus or the representation precision is determined based on the information representing the VTI representation precision in the video code string.

[0091] In FIGS. 15A to 15E, the bit width of each field is limited for describing the time stamp formats, but each bit width may be previously determined in response to the application and is not limited to the bit widths shown in the figures. The origin of the time represented by the time stamp need not necessarily begin with zero and may be selected at random for improving safety if the communication line is encrypted. (Fifth embodiment)

[0092] A fifth embodiment of the invention is the same as the second and third embodiments in the basic configurations of coding apparatus and decoding apparatus; they differ only in M bit field added to an RTP header and therefore only the differences will be discussed in detail.

[0093] The M bit (M in FIG. 10) is a one-bit flag contained in an RTP header indicating that such information for causing a particularly important event to occur is contained in one packet as compared with any other packet; it is previously determined in response to the type of multimedia information put on RTP payload. The M bit may be set, for example, as follows:

(1) M is set to 1 only for the RTP packet containing a GOV header and the RTP packet containing a VOP header of VOP (I-VOP) undergoing intraframe coding; M is set to 0 for other RTP packets.
(2) M is set to 1 only for the last RTP packet if one VOP head is divided across RTP packets.
(3) M is set to 1 only if more than one VOP head is contained in an RTP packet.
(4) M is set to 1 only if more than one video packet is contained in an RTP packet.
(5) M is set to 1 only if RTP payload begins at the top of each layer shown in FIG. 2.

[0094] To define the M bit as in (1), the advantage is provided that the fact that the packet with the M bit set to 1 is a packet containing video information that can become a random access point can be easily known. That is, in other methods, unless the header information of MPEG4 video code bit string contained in RTP payload is decoded, whether or not it is a random access point cannot be determined; however, in the method, processing of the RTP header process portion in a communication unit on a transmission line or in the receiving party is only performed, whereby whether or not the current packet being processed contains information that can become a random access point is known, and processing is very facilitated in searching for a random access point.

[0095] To define the M bit as in (2), whether or not transmission of one VOP is complete can be determined based the M bit in such a case where VOP is divided across RTP packets and transmitted if the packet length of RTP payload is short as compared with the number of code bits of VOP, usually observed when the code bit rate is high. This has a good affinity for definition of the RTP format for MPEG1/MPEG2 video shown in document 4, and commonality of processing can be easily accomplished.

[0096] In contrast, the definition of the M bit in (3) or (4) indicating that more than one VOP or video packet is contained in one RTP packet has effectiveness in such

BNSDOCID: <EP_____1035735A2_I_>

a case where the packet length of RTP payload is equal to or comparatively longer than the code bit length of VOP in such application where the code bit rate is comparatively low.

[0097] To define the M bit as in (5), whether or not the header information of each layer in MPEG4 video code string is contained in the RTP packet is indicated, and the definition of the M bit becomes effective for protecting the important information contained in the header information. As the header types, more particularly, configuration information functions (VisualObjectSequence(), VisualObject(), VisualObjectLayer(), or entry point functions for elementary streams (Group of_VideoObjectPlane(), VideoObjectPlane(), video plane_with_short_header(), MeshObject(), FaceObject()) are included.

(Sixth embodiment)

[0098] A sixth embodiment of the invention is the same as the first embodiment in the basic configurations of coding apparatus and decoding apparatus; they differ only in dividing rules of video code string in access unit generators 31a to 31e and sync layer packet generator and therefore only the differences will be discussed in detail.

[0099] When a sync layer packet is divided and put on RTP payload, satisfying all the following four items may be adopted as a rule:

(3-1) Each header above the VOL in the hierarchical structure in FIG. 2 must be placed at the top of the sync layer packet payload (just after the sync layer packet header) or just after the higher-layer header;
(3-2) a higher-layer header than the header placed at the top of the sync layer packet payload must not exist at an intermediate point of the payload;
(3-3) if one or more headers exist in the sync layer packet payload, the payload must always begin with the header; and
(3-4) header must not be divided across sync layer packets.

[0100] These differ from the dividing rules (1-1) to (1-4) shown in the first embodiment only in handling the GOV header.

(Seventh embodiment)

[0101] A seventh embodiment of the invention is the same as the second and third embodiments in the basic configurations of coding apparatus and decoding apparatus; they differ only in dividing rules of video code string put on RTP payload and therefore only the differences will be discussed in detail.

[0102] When a video code string is divided and put on RTP payload, satisfying all the following four items may be adopted as a rule:

(4-1) Each header above the VOL in the hierarchical structure in FIG. 2 must be placed at the top of the RTP payload (just after the RTP header) or just after the higher-layer header;
(4-2) a higher-layer header than the header placed at the top of the RTP payload must not exist at an intermediate point of tire payload;
(4-3) if one or more headers exist in the RTP payload, the payload must always begin with the header; and
(4-4) video header must not be divided across RTP packets.

[0103] These differ from the dividing rules (2-1) to (2-4) shown in the second embodiment only in handling the GOV header.

[0104] FIGS. 17A and 17C are drawings to describe RTP packet division prohibited in the rules (4-1) to (4-4); FIGS. 17A and 17C show examples of RTP packets not prepared if RTP packet division is executed according to the rules, whereas FIG. 17B shows an example prepared based on the rule.

[0105] In FIG. 17A, a VOP header is divided across RTP packets, but dividing the video header across RTP packets is prohibited based on the rule (4-4). A VOP start code is prefixed to the top of the VOP header and the decoder can determine the top position of the VOP header based on the start code. However, if the VOP header is divided as shown in FIG. 17A, no VOP start code exists in the second RTP packet. Thus, if the first RTP packet in the figure is lost, the top position of the VOP header is not found, making it impossible for the decoder to decode the VOP header correctly. Thus, dividing the video header across RTP packets is prohibited according to the division rule. FIG. 17A shows the VOP header example, but the description also applies to any other video header, such as a VS header, a VO header, a VOL header, or a video packet header.

[0106] FIGS. 17B and 17C show examples wherein two video packets are divided in two RTP packets. FIG. 17C shows an example of violating the division rule (4-3) because video packet header (VP header) is placed at a position other than the top of RTP payload in the second RTP packet.

[0107] In FIG. 17B, one video packet is entered in one RTP packet; in FIG. 17C, the first video packet is divided across two RTP packets and the latter half of the first video packet is entered in the same RTP packet as the second video packet. If RTP packet division is executed corresponding to video packet as shown in FIG. 17B, even if one RTP packet is lost due to an error, the video packet entered in the other RTP packet can be decoded. In contrast, in FIG. 17C, if the second RTP packet is lost, information not only in the second video packet, but also in the first video packet is lost, thus both video packets cannot be decoded correctly. Therefore,

BNSDOCID: <EP_____1035735A2_I_>

dividing as in FIG. 17C is prohibited according to the division rule.

[0108] The RTP packet division examples prohibited according to the division rules (4-1) to (4-4) have been described; if the division rules (2-1) to (2-4) are used, RTP packet preparation as in FIGS. 17A and 17C are also prohibited.

[0109] Next, a specific example of information storage media according to the invention will be discussed.

[0110] FIG. 18 is a block diagram to show a system for using a coding apparatus to prepare RTP and record it on a record medium according to the invention. Numeral 880 denotes a video signal input unit for inputting a video signal. The video signal input unit is, for example, a video camera. Alternatively, a video signal recorded on a record medium (not shown) may be input or a video signal may be input from another apparatus or system via a transmission line (not shown). A video coder 870 performs moving image coding on an input video signal 852 and outputs a video code string 857. The video code string 857 is input to an RTP transmitter 855, which then outputs an RTP packet 851. The RTP packet 851 is recorded on a storage medium 860. Information indicating the length of RTP packet (not shown) may also be recorded on a record medium 810.

[0111] FIG. 19 is a block diagram to show a system for reproducing a video signal using the record medium 810 prepared using the system in FIG. 18. A code string containing an RTP packet coded by the coding apparatus according to the invention is stored on the record medium 810. Numeral 805 denotes an RTP receiver for decoding an RTP packet 801 recorded on the record medium 810. The RTP receiver 805 decodes the time stamp and the sequence number of an RTP packet header and outputs a video code string 807. If information indicating the length of RTP packet (not shown) is also recorded on the record medium 810, the information is also input to the RTP receiver 805 for executing RTP decoding. Numeral 820 denotes a video decoder for reproducing a video playback signal 802 from the video code string 807. Numeral 830 denotes a video signal output unit for outputting a video signal. The video signal output unit is, for example, a display. Alternatively, a reproduced video signal may be recorded on a storage medium (not shown) or may be transmitted to another apparatus or system via a transmission line (not shown).

[0112] The described system stores RTP packets in the format previously covered in the description of the embodiments on the storage medium 810. The RTP packets are characterized by the fact that RTP packet division is executed based on the RTP packet division rules (1-1) to (1-4), (2-1) to (2-4), and (4-1) to (4-4) and that the time stamp of each RTP header is prepared by converting the bit format of the time stamp of the video code string as described above.

[0113] In the example in FIG. 18, in the whole system, only one video playback signal is input and one video coder and one RTP transmitter prepare an RTP packet. However, as in the above-described embodiments, more than one RTP transmitter and more than one video coder may be used to code more than one video signal. In this case, a plurality of RTP packet strings corresponding to a plurality of video input signals may be stored on the storage medium 860 or separate storage media may be used in one-to-one correspondence with the video playback signals.

[0114] In the example in FIG. 19, the whole system contains one RTP receiver and one video decoder and reproduces only one video playback signal. However, as in the above-described embodiments, more than one RTP receiver and more than one video decoder may be used to reproduce more than one video playback signal. In this case, a plurality of RTP packet strings corresponding to a plurality of video playback signals may be recorded on the record medium 810 or separate storage media may be used in one-to-one correspondence with the video playback signals. A plurality of video playback signals may be output to separate video signal output units or a plurality of video signals may be combined by a video signal combiner (not shown) and output to one video signal output unit.

[0115] FIG. 20 is a flowchart to show processing of executing moving image coding and RTP packet preparation and recording the RTP packets on the storage medium in the coding system in FIG. 18.

[0116] First, the video coder 870 prepares a video initial header and outputs it to the RTP transmitter 855 at step S01. The video initial header corresponds to the VS, VO, VOL header in the video syntax structure previously described with reference to FIG. 2, for example, and indicates the coding mode of one whole video stream. Next, an RTP header is initialized at step S02. In the RTP header, the payload type (PT) and SSRC, each an information piece taking a given value for one video input signal, are set. The initial values of the sequence number (SN) and the time stamp are also set. The initial values of the sequence number (SN) and the time stamp may be set to fixed values (for example, 0) or may be random numbers. Next, with the video initial header prepared at step S01 as RTP payload; the initial RTP header prepared at step S02 is added and an initial RTP packet is prepared at step S03. Further, the prepared initial RTP packet is recorded on the storage medium 860 at step S04.

[0117] At steps S05 to S17, a video signal is input one frame (VOP, also called a picture) at a time, moving image coding is performed, and an RTP packet is prepared and recorded. First, one frame of a video signal is input from the video signal input unit 880 at step S05. The video coder 870 converts one frame of the video signal input into a moving image code string at step S06. The time stamp of the RTP header is calculated at step S07. The time stamp may be calculated based on time stamp information modulo_time_base (MTB) and VOP_time_increment (VTI) of video code string as pre-

BNSDOCID: <EP_____1035735A2_I_>

viously described in the embodiment.

[0118] The moving image code string provided at step S06 is output one video packet at a time and is input to the RTP transmitter 855 at step S08. At steps S08 to S16, the RTP transmitter 855 prepares and records an RTP packet while inputting one video packet at a time.

[0119] At steps S09 to S11, the marker bit (M) of the RTP header is calculated. Whether or not the input video packet is the last video packet in one frame is determined at step S09. If the video packet is the last video packet, M is set to 1 at step S10; otherwise, M is set to 0 at step S11.

[0120] Next, padding processing of the RTP payload is performed and the padding flag bit (P) of the RTP header is set at step S12. The length of the input video packet is calculated and if the length is a multiple of 32 bits, the padding flag (P) of the RTP header is set to 0 and the video packet is used as RTP payload intact. If the length is not a multiple of 32 bits, the padding flag is set to 1 and padding bits are added to the tail of the video packet so that the length of RTP load becomes a multiple of 32 bits.

[0121] In the RTP header, as information other than the marker bit or the padding flag set at steps S09 to S12, the values set at other steps are used. The thus setup RTP header and RTP payload are combined to prepare an RTP packet at step S13. The prepared RTP packet is recorded on the storage medium 860 at step S14. Whenever one RTP packet is generated and recorded, the sequence number (SN) is incremented by one at step S15. Next, whether the marker bit of the RTP header is 0 or 1 is determined at step S16 and branch processing is performed as follows: If M=0, the processed video packet is not the last video packet in the frame. Then, control returns to step S08 for repeating processing of inputting one video packet at a time and preparing and recording an RTP packet. If M=1, the processed video packet is the last video packet in the frame. Then, control goes to step S17. At step S17, whether or not the processed frame is the last frame of the video signal is determined. If the processed frame is the last frame, termination processing is performed. If the processed frame is not the last frame, control returns to step S05 for repeating processing of inputting the video signal one frame at a time, performing moving image coding, and preparing and recording an RTP packet.

[0122] In FIG. 18, numerals 861 to 863 indicate examples of RTP packets prepared and recorded according to the flowchart of FIG. 20. Numeral 861 indicates an example of an initial RTP packet prepared and recorded at steps S01 to S04. Numerals 862 and 863 indicate examples of RTP packets prepared and recorded at steps S05 to S17.

[0123] Next, as an application example of the invention, an embodiment of a moving image transmission system incorporating the coding apparatus and the decoding apparatus of the invention will be discussed with reference to FIG. 12.

[0124] A moving image signal input from a camera (not shown) installed in a personal computer 1001 undergoes moving image coding and RTP coding performed by the coding apparatus (or coding software) built in the personal computer 1001. An RTP packet output from the coding apparatus is transmitted by wireless by a radio 1003 together with any other voice and data information, and is received by another radio 1004. For example, portable telephones, PHSs, wireless LAN units, etc., may be used as the radios. The signal received at the radio 1004 is disassembled into the RTP packet of the moving image signal and the voice and data information. The RTP packet of the moving image signal is decoded by the decoding apparatus (or decoding software) built in a notebook computer 1005 and is displayed on a display of the notebook computer 1005. On the other hand, a moving image signal input from a camera (not shown) installed in the notebook computer 1005 is coded in a similar manner to that described above using the coding apparatus (or coding software) built in the notebook computer 1005. A prepared RTP packet and any other voice and data information are multiplexed and transmitted by wireless by the radio 1004 and received by the radio 1003. The signal received by the radio 1003 is disassembled into the RTP packet of the moving image signal and the voice and data information. The RTP packet of the moving image signal is decoded by the decoding apparatus (or decoding software) built in the personal computer 1001 and is displayed on a display of the personal computer 1001.

[0125] The coding apparatus and the decoding apparatus according to the invention can also be applied to moving image communication between the personal computer 1001 or the notebook computer 1005 and a portable videophone 1006. An RTP packet prepared by the coding apparatus built in the personal computer 1001 or the notebook computer 1005 and transmitted by wireless by the radio 1003 or 1004 is received at a radio built in the portable videophone 1006. The signal received at the radio is disassembled into the RTP packet of the moving image signal and the voice and data information. The RTP packet of the moving image signal is decoded by the decoding apparatus (or decoding software) built in the portable videophone 1006 and is displayed on a display of the portable videophone 1006. On the other hand, a moving image signal input from a camera 1007 built in the portable videophone 1006 is coded in a similar manner to that in the examples of the personal computer 1001 and the notebook computer 1005 described above using the coding apparatus (or coding software) built in the portable videophone 1006. A prepared RTP packet and any other voice and data information are multiplexed and transmitted by wireless by the radio built in the portable videophone 1006 and received by the radio 1003 or 1004. The signal received by the radio 1003 or 1004 is

BNSDOCID: <EP_____1035735A2_I_>

disassembled into the RTP packet of the moving image signal and the voice and data information. The RTP packet of the moving image signal is decoded by the decoding apparatus (or decoding software) built in the personal computer 1001 or the notebook computer 1005 and is displayed on the display of the personal computer 1001 or the notebook computer 1005.

[0126] As described throughout the specification, according to the invention, to divide a video code string provided by compressing and coding a video signal and enter in an RTP packet for transmission, the above-described dividing rules are used to enter header information in the video code string in the top of a sync layer packet or RTP payload, whereby the duplication function of important information provided by video coding is used effectively and resistance to a packet loss of RTP packet can be enhanced.

Claims

1. A moving image coding apparatus comprising:

   coding means for dividing an input moving image signal into a plurality of frame image signals, dividing each of the frame image signals into one or more area image signals, and compression coding the area image signal into an area image code string, and adding a frame header information indicating a coding mode of the frame to the area image code string; and
   packetization means for collecting one or more area image code strings to which the frame header information is added, and adding packet header information.

2. The moving image coding apparatus as claimed in claim 1 wherein said packetization means includes a multiplexer comprising a plurality of access unit generators for separating the code strings into predetermined units and generating access units and a sync layer packet generator for receiving the access units from the access unit generators and generating a sync layer packet.

3. A moving image decoding apparatus comprising:

   reception means for receiving a moving image code string put into a packet;
   separation means for separating one or more area image code strings contained in each packet of the moving image code string;
   area image decoding means for decoding the separated area image code string and outputting a decoded area image signal;
   image frame decoding means for assembling the decoded area image signal for each frame and outputting a decoded frame image signal; and
   means for generating a decoded moving image signal based on the decoded frame image signal.

4. The moving image decoding apparatus as claimed in claim 3 wherein said separation means comprises a decoder for decoding an access unit based on information of a sync layer packet header contained in the input code string and an access unit decoder for decoding an access unit header and generating an original code string.

5. A moving image coding apparatus comprising:

   a plurality of coding means for dividing an input moving image signal into a plurality of frame image signals, dividing each of the frame image signals into one or more area image signals, and compression coding the area image signal into an area image code string, and adding a frame header information indicating a coding mode of the frame to the area image code string; and
   a plurality of packetization means for collecting one or more area image code strings to which the frame header information is added, and adding packet header information.

6. A moving image decoding apparatus comprising:

   a plurality of reception means for receiving a plurality of moving image code strings put into packet;
   area image decoding means for decoding area image code strings of the moving image code strings input from said plurality of reception means and outputting a plurality of decoded area image signals;
   frame image decoding means for assembling the decoded area image signals for each frame and outputting a decoded image frame signal; and
   means for generating a decoded moving image signal based on the decoded image frame signal.

7. The moving image coding apparatus as claimed in claim 1 wherein said packet header information includes time stamp information generated by converting time stamp information in the code strings into a predetermined format.

8. The moving image coding apparatus as claimed in claim 6 wherein said packet header information includes time stamp information generated by converting time stamp information in the code strings into a predetermined format.

BNSDOCID: <EP_____1035735A2_I_>

9. The moving image decoding apparatus as claimed in claim 6 wherein said reception means has means for restoring time stamp information of an image contained in packet header information to the original from a predetermined format in said area image decoding means and said frame image decoding means.

10. A record medium recording a code string prepared by a moving image coding apparatus comprising: coding means for dividing an input moving image signal into a plurality of frame image signals, dividing each of the frame image signals into one or more area image signals, and compression coding the area image signal into an area image code string, and adding a frame header information indicating a coding mode of the frame to the area image code string; and packetization means for collecting one or more area image code strings to which the frame header information is added, and adding packet header information.

11. The moving image coding apparatus as claimed in claim 5 wherein said packetization means includes a multiplexer comprising a plurality of access unit generators for separating the code strings into predetermined units and generating access units and a sync layer packet generator for receiving the access units from the access unit generators and generating a sync layer packet.

12. The moving image decoding apparatus as claimed in claim 6 wherein said separation means comprises a decoder for decoding an access unit based on information of a sync layer packet header contained in the input code string and an access unit decoder for decoding an access unit header and generating an original code string.

13. The moving image coding apparatus as claimed in claim 9 wherein said packet header information includes time stamp information generated by converting time stamp information in the code strings into a predetermined format.

14. A method of coding a moving image, comprising the steps of:

dividing an input moving image signal into a plurality of frame image signals;
dividing each of the frame image signals into one or more area image signals;
compression coding the area image signal into an area image code string;
adding a frame header information indicating a coding mode of the frame to the area image code string; and
collecting one or more area image code strings to which the frame header information is added, and adding packet header information.

15. The method of coding a moving image as claimed in claim 14, further comprising the steps of:

separating the code strings into predetermined units and generating access units; and
receiving the access units from the access unit generators and generating a sync layer packet.

16. A method of coding a moving image, comprising the steps of:

dividing an input moving image signal into a plurality of frame image signals;
dividing each of the frame image signal into one or more area image signals;
compression coding the area image signal into an area image code string;
adding a frame header information indicating a coding mode of the frame to the area image code string; and
collecting one or more area image code strings to which the frame header information is added, and adding packet header information.

17. A recording medium for executing computer program comprising the steps of:

dividing an input moving image signal into a plurality of frame image signals;
dividing each of the frame image signals into one or more area image signals;
compression coding the area image signal into an area image code string;
adding a frame header information indicating a coding mode of the frame to the area image code string; and
collecting one or more area image code strings to which the frame header information is added, and adding packet header information.

18. The recording medium for executing computer program as claimed in claim 17, wherein said computer program further comprising the steps of:

separating the code strings into predetermined units and generating access units; and
receiving the access units from the access unit generators and generating a sync layer packet.

19. A recording medium for executing computer program comprising the steps of:

dividing an input moving image signal into a plurality of frame image signals;
dividing each of the frame image signal into

BNSDOCID: <EP_____1035735A2_I_>

one or more area image signals;
compression coding the area image signal into an area image code string;
adding a frame header information indicating a coding mode of the frame to the area image code string; and
collecting one or more area image code strings to which the frame header information is added, and adding packet header information.

20. A method of decoding a moving image, comprising the steps of:

receiving a moving image code string put into a packet;
separating one or more area image code strings contained in each packet of the moving image code string;
decoding the separated area image code string and outputting a decoded area image signal;
assembling the decoded area image signal for each frame and outputting a decoded frame image signal; and
generating a decoded moving image signal based on the decoded frame image signal.

21. The method of decoding a moving image as claimed in claim 20, further comprising the steps of:

decoding an access unit based on information of a sync layer packet header contained in the input code string; and
decoding an access unit header and generating an original code string.

22. A method of decoding a moving image, comprising the steps of:

receiving a plurality of moving image code strings put into packet;
decoding area image code strings of the moving image code strings input from said plurality of reception means and outputting a plurality of decoded area image signals;
assembling the decoded area image signals for each frame and outputting a decoded image frame signal; and
generating a decoded moving image signal based on the decoded image frame signal.

23. A recording medium for executing computer program comprising the steps of:

receiving a moving image code string put into a packet;
separating one or more area image code strings contained in each packet of the moving image code string;
decoding the separated area image code string and outputting a decoded area image signal;
assembling the decoded area image signal for each frame and outputting a decoded frame image signal; and
generating a decoded moving image signal based on the decoded frame image signal.

24. The recording medium for executing computer program as claimed in claim 22, wherein said computer program further comprises the steps of:

decoding an access unit based on information of a sync layer packet header contained in the input code string; and
decoding an access unit header and generating an original code string.

25. A recording medium for executing computer program comprising the steps of:

receiving a plurality of moving image code strings put into packet;
decoding area image code strings of the moving image code strings input from said plurality of reception means and outputting a plurality of decoded area image signals;
assembling the decoded area image signals for each frame and outputting a decoded image frame signal; and
generating a decoded moving image signal based on the decoded image frame signal.

26. The moving image coding apparatus as claimed in claim 1, wherein said frame header information includes any information of a time-code, a VPO coding mode, intra DC VLC table change information, motion vector range information contained in the VOP header.

27. The moving image coding apparatus as claimed in claim 3, wherein said frame header information includes any information of a time code, a VPO coding mode, intra DC VLC table change information, motion vector range information contained in the VOP header.

28. The moving image coding apparatus as claimed in claim 5, wherein said frame header information includes any information of a time code, a VPO coding mode, intra DC VLC table change information, motion vector range information contained in the VOP header.

29. The moving image coding apparatus as claimed in claim 6, wherein said frame header information includes any information of a time code, a VPO coding mode, intra DC VLC table change informa-

BNSDOCID: <EP_____1035735A2_I_>

tion, motion vector range information contained in the VOP header.

30. The moving image coding apparatus as claimed in claim 9, wherein said frame header information includes any information of a time code, a VPO coding mode, intra DC VLC table change information, motion vector range information contained in the VOP header.

31. A moving image coding apparatus comprising:

a coder configured to perform a function for dividing an input moving image signal into a plurality of frame image signals, dividing each of the frame image signals into one or more area image signals, and compression coding the area image signal into an area image code string, and adding a frame header information indicating a coding mode of the frame to the area image code string; and
a packetizator configured to perform a function for collecting one or more area image code strings to which the frame header information is added, and adding packet header information.

32. The moving image coding apparatus as claimed in claim 31 wherein said packetizatior includes a multiplexer comprising a plurality of access unit generators configured to perform a function for separating the code strings into predetermined units and generating access units and a sync layer packet generator for receiving the access units from the access unit generators and generating a sync layer packet.

33. A moving image decoding apparatus comprising:

a receiver configured to perform a function for receiving a moving image code string put into a packet;
a separator for separating one or more area image code strings contained in each packet of the moving image code string;
an area image decoder configured to perform a function for decoding the separated area image code string and outputting a decoded area image signal;
an image frame decoder to perform a function for assembling the decoded area image signal for each frame and outputting a decoded frame image signal; and
a generator configured to perform a function for generating a decoded moving image signal based on the decoded frame image signal.

34. The moving image decoding apparatus as claimed in claim 33 wherein said separator comprises a decoder configured to perform a function for decoding an access unit based on information of a sync layer packet header contained in the input code string and an access unit decoder configured to perform a function for decoding an access unit header and generating an original code string.

35. A moving image coding apparatus comprising:

a plurality of coders configured to perform a function for dividing an input moving image signal into a plurality of frame image signals, dividing each of the frame image signals into one or more area image signals, and compression coding the area image signal into an area image code string, and adding a frame header information indicating a coding mode of the frame to the area image code string; and
a plurality of packetizators configured to perform a function for collecting one or more area image code strings to which the frame header information is added, and adding packet header information.

36. A moving image decoding apparatus comprising:

a plurality of receivers configured to perform a function for receiving a plurality of moving image code strings put into packet;
an area image decoder configured to perform a function for decoding area image code strings of the moving image code strings input from said plurality of receivers and outputting a plurality of decoded area image signals;
a frame image decoder configured to perform a function for assembling the decoded area image signals for each frame and outputting a decoded image frame signal; and
a generator configured to perform a function for generating a decoded moving image signal based on the decoded image frame signal.

37. The moving image coding apparatus as claimed in claim 31 wherein said packet header information includes time stamp information generated by converting time stamp intonation in the code strings into a predetermined format.

38. The moving image coding apparatus as claimed in claim 36 wherein said packet header information includes time stamp information generated by converting time stamp information in the code strings into a predetermined format.

39. The moving image decoding apparatus as claimed in claim 36 wherein said receiver has a unit configured to perform a function for restoring time stamp information of an image contained in packet header

BNSDOCID: <EP_____1035735A2_I_>

information to the original from a predetermined format in said area image decoder and said frame image decoder.

40. The moving image coding apparatus as claimed in claim 35 wherein said packetizator includes a multiplexer comprising a plurality of access unit generators configured to perform a function for separating the code strings into predetermined units and generating access units and a sync layer packet generator configured to perform a function for receiving the access units from the access unit generators and generating a sync layer packet.

41. The moving image decoding apparatus as claimed in claim 36 wherein said separator comprises a decoder configured to perform a function for decoding an access unit based on information of a sync layer packet header contained in the input code string and an access unit decoder configured to perform a function for decoding an access unit header and generating an original code string.

42. The moving image coding apparatus as claimed in claim 39 wherein said packet header information includes time stamp information generated by converting time stamp information in the code strings into a predetermined format.

43. The moving image coding apparatus as claimed in claim 31, wherein said frame header information includes any information of a time code, a VPO coding mode, intra DC VLC table change information, motion vector range information contained in the VOP header.

44. The moving image coding apparatus as claimed in claim 33, wherein said frame header information includes any information of a time code, a VPO coding mode, intra DC VLC table change information, motion vector range information contained in the VOP header.

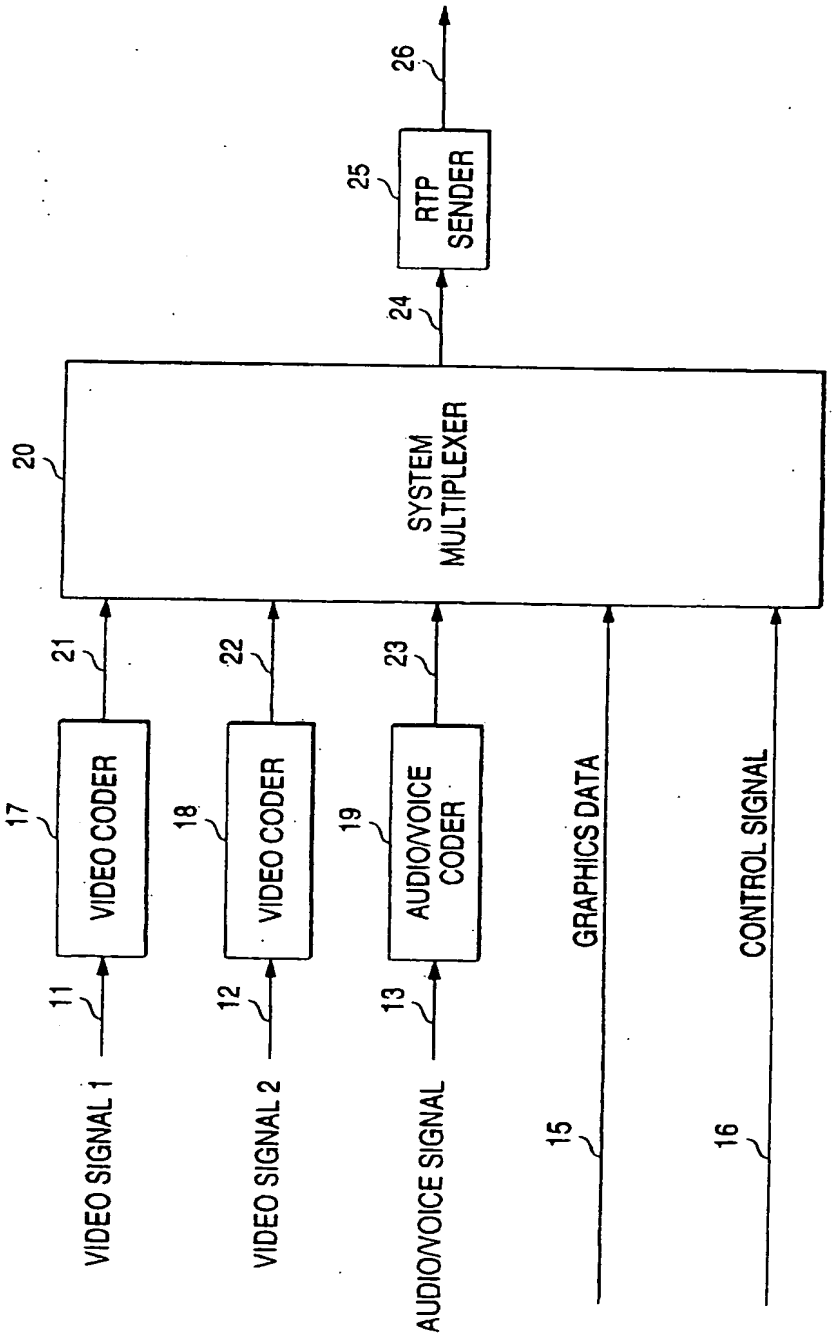
45. The moving image coding apparatus as claimed in claim 35, wherein said frame header information includes any information of a time code, a VPO coding mode, intra DC VLC table change information, motion vector range information contained in the VOP header.

46. The moving image coding apparatus as claimed in claim 36, wherein said frame header information includes any information of a time code, a VPO coding mode, intra DC VLC table change information, motion Vector range information contained in the VOP header.

47. The moving image coding apparatus as claimed in claim 39,wherein said frame header information includes any information of a time code, a VPO coding mode, intra DC VLC table change information, motion vector range information contained in the VOP header.

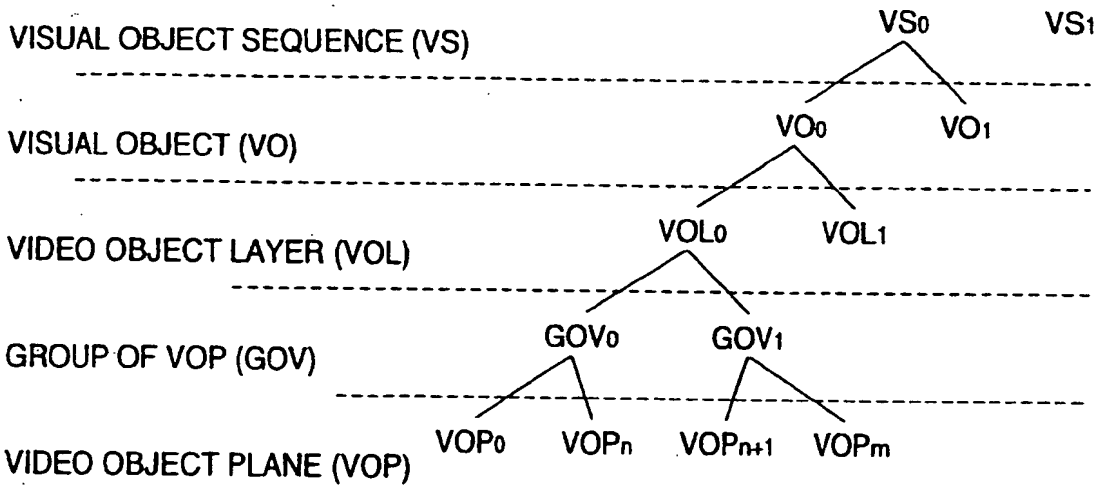
BNSDOCID: <EP_____1035735A2_I_>

FIG. 1



BNSDOCID: <EP_____1035735A2_I_>

FIG. 2



BNSDOCID: <EP_____1035735A2_I_>

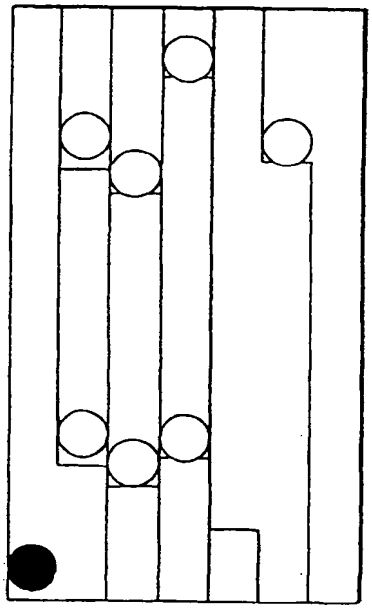


FIG. 3A

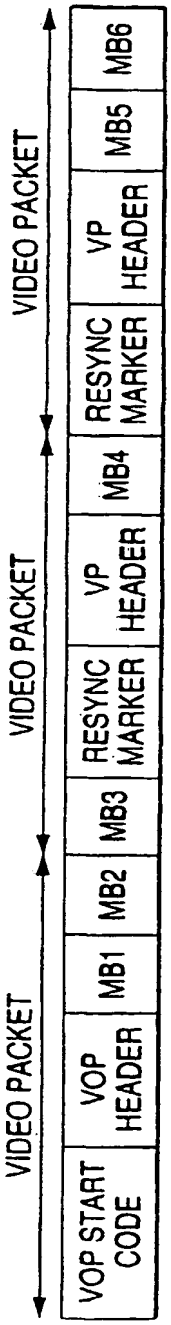


FIG. 3B

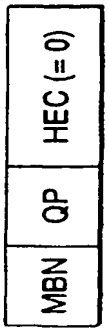


FIG. 3C

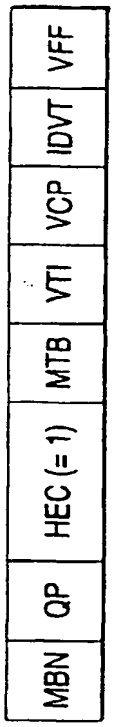
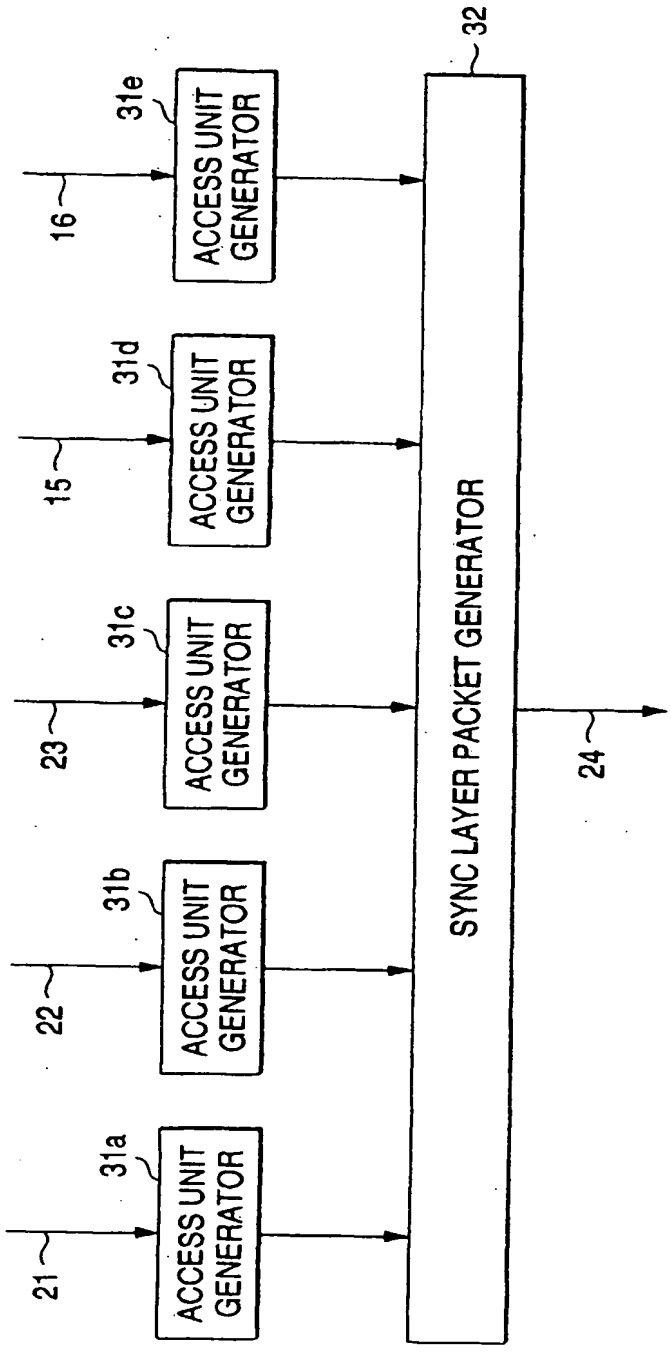


FIG. 3D

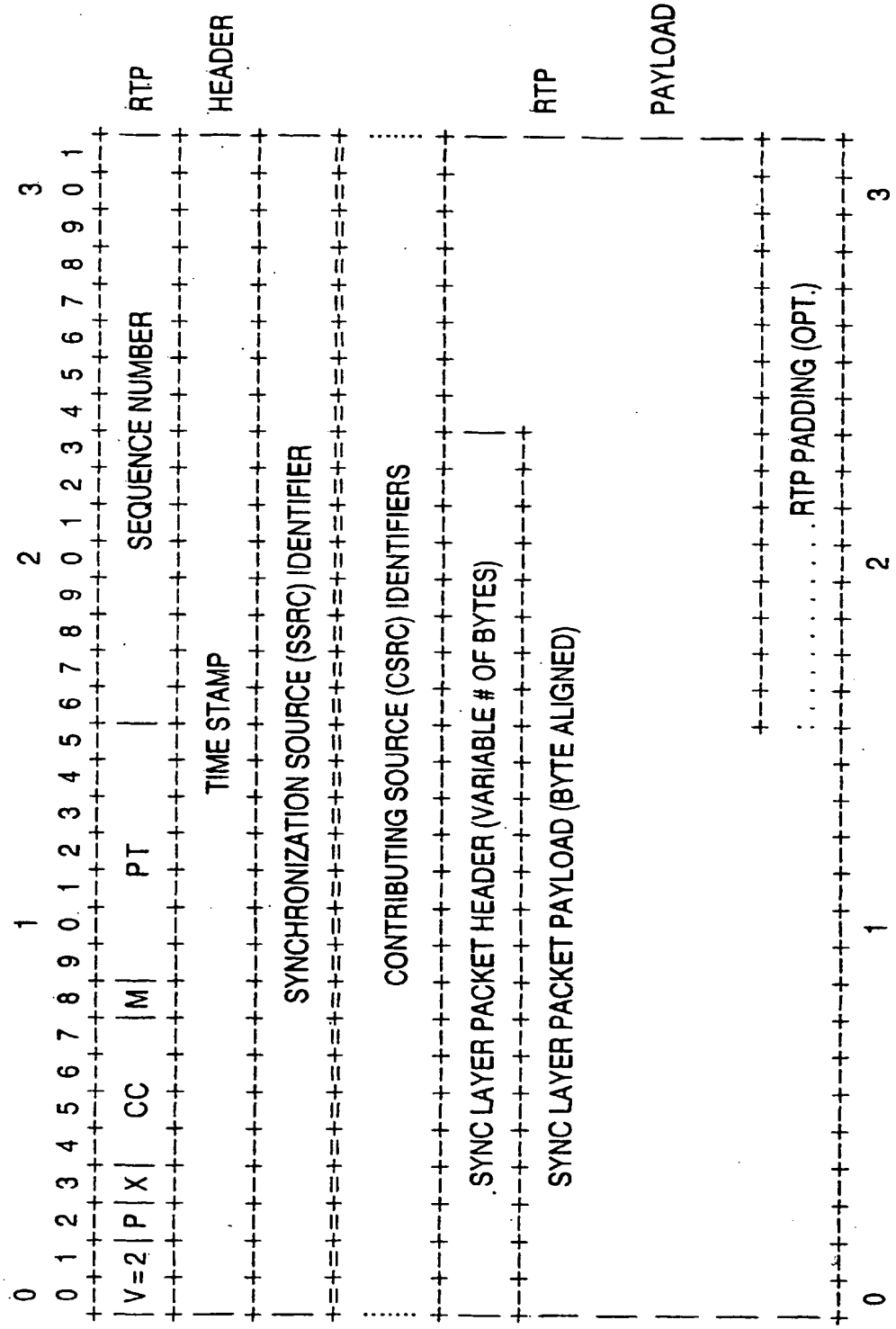
BNSDOCID: <EP_____1035735A2_I_>

FIG. 4



BNSDOCID: <EP_____1035735A2_I_>

FIG. 5



BNSDOCID: <EP_____1035735A2_I_>

FIG. 6A

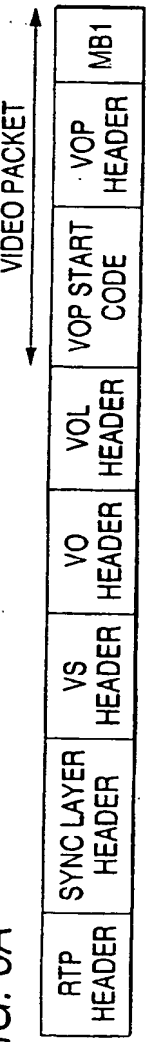


FIG. 6B

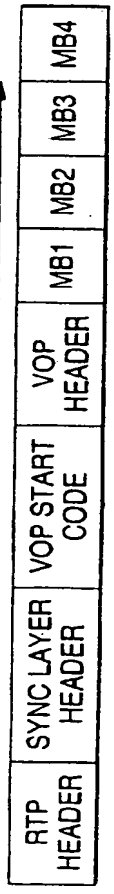


FIG. 6C

FIG. 6D

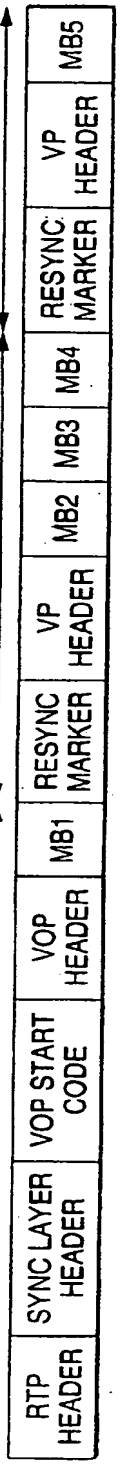
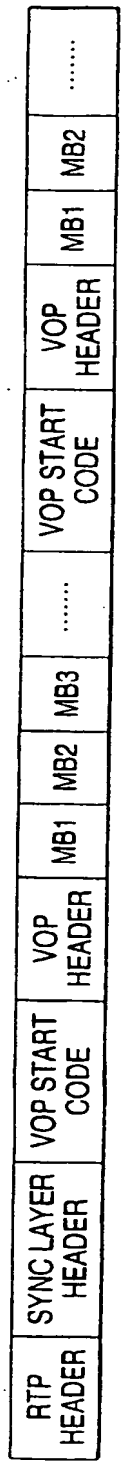
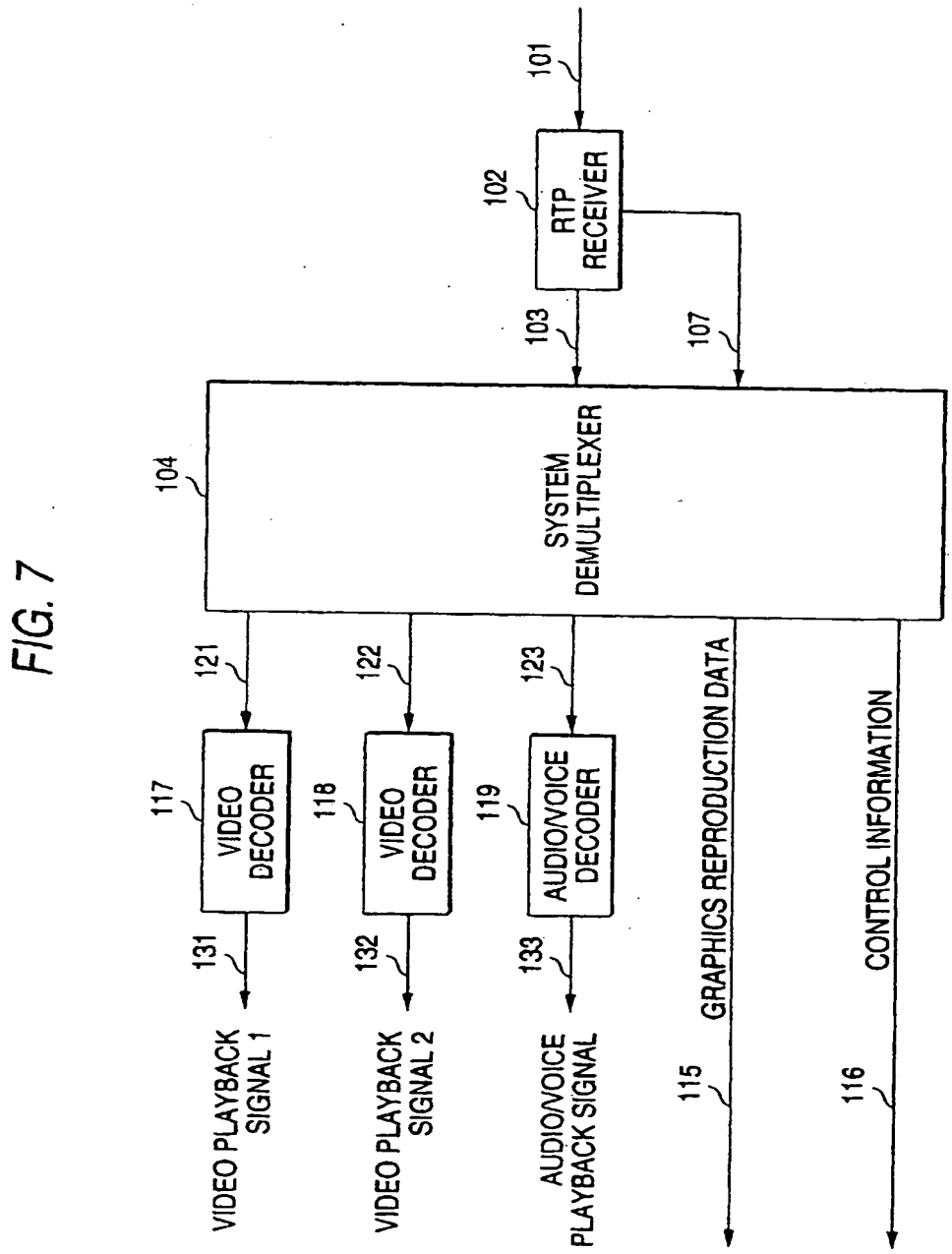


FIG. 6E



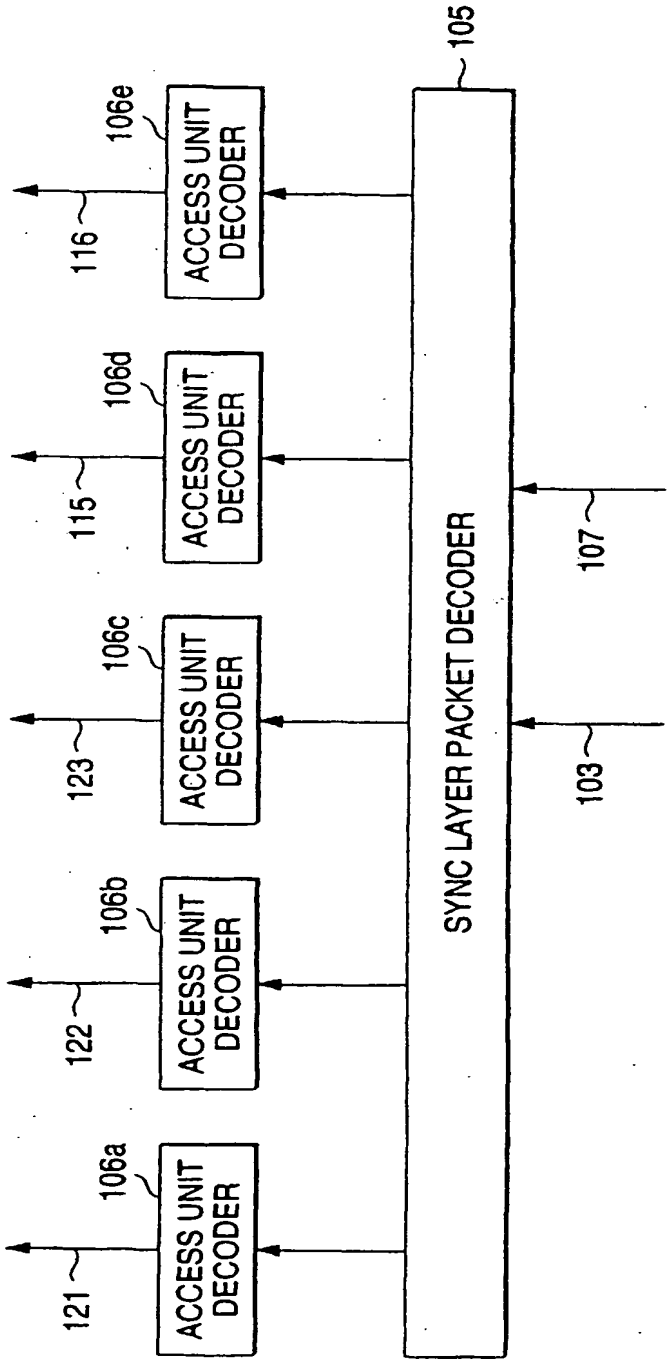
BNSDOCID: <EP_____1035735A2_I_>

FIG. 7



BNSDOCID: <EP_____1035735A2_I_>

FIG. 8



BNSDOCID: <EP_____1035735A2_I_>

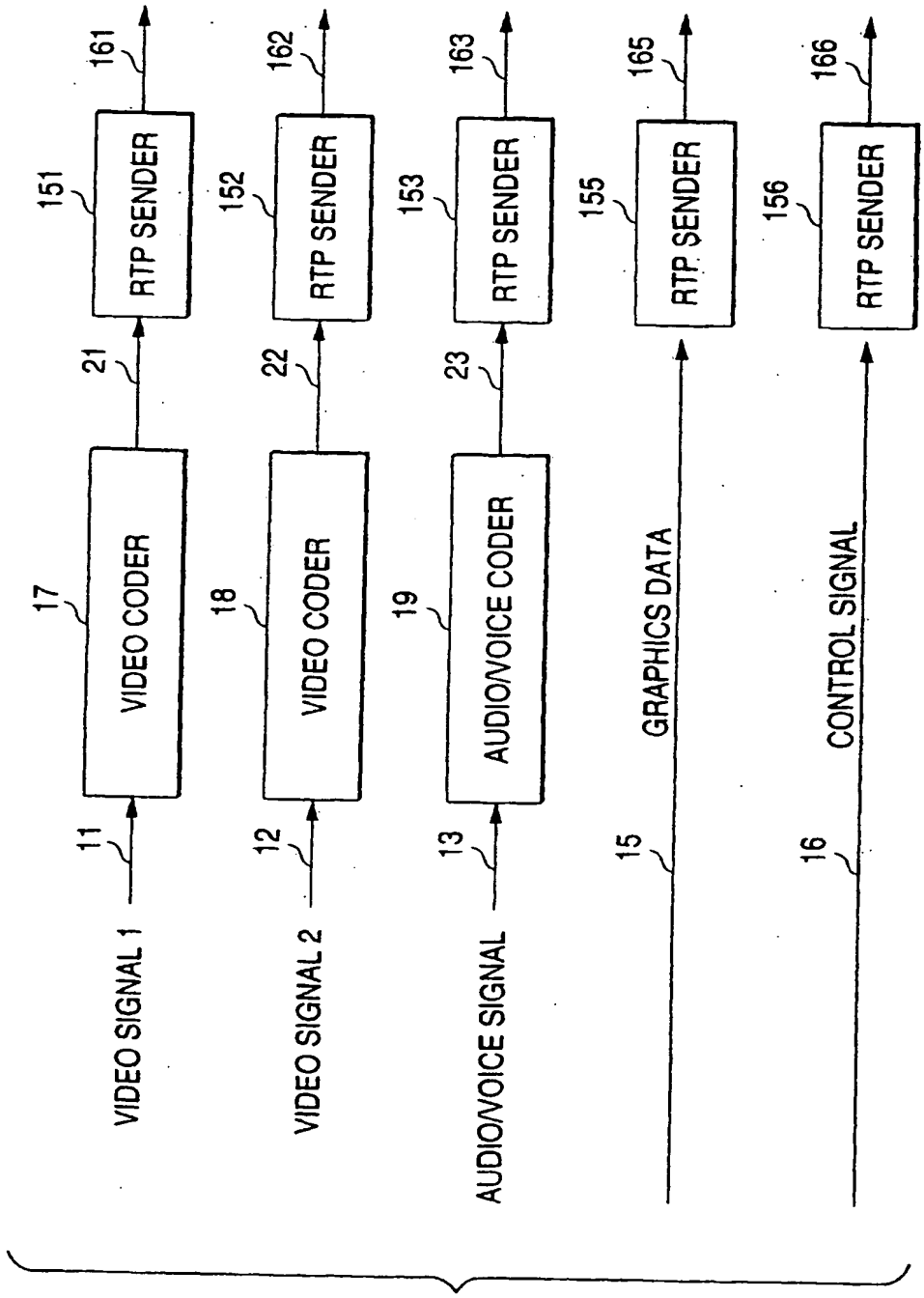
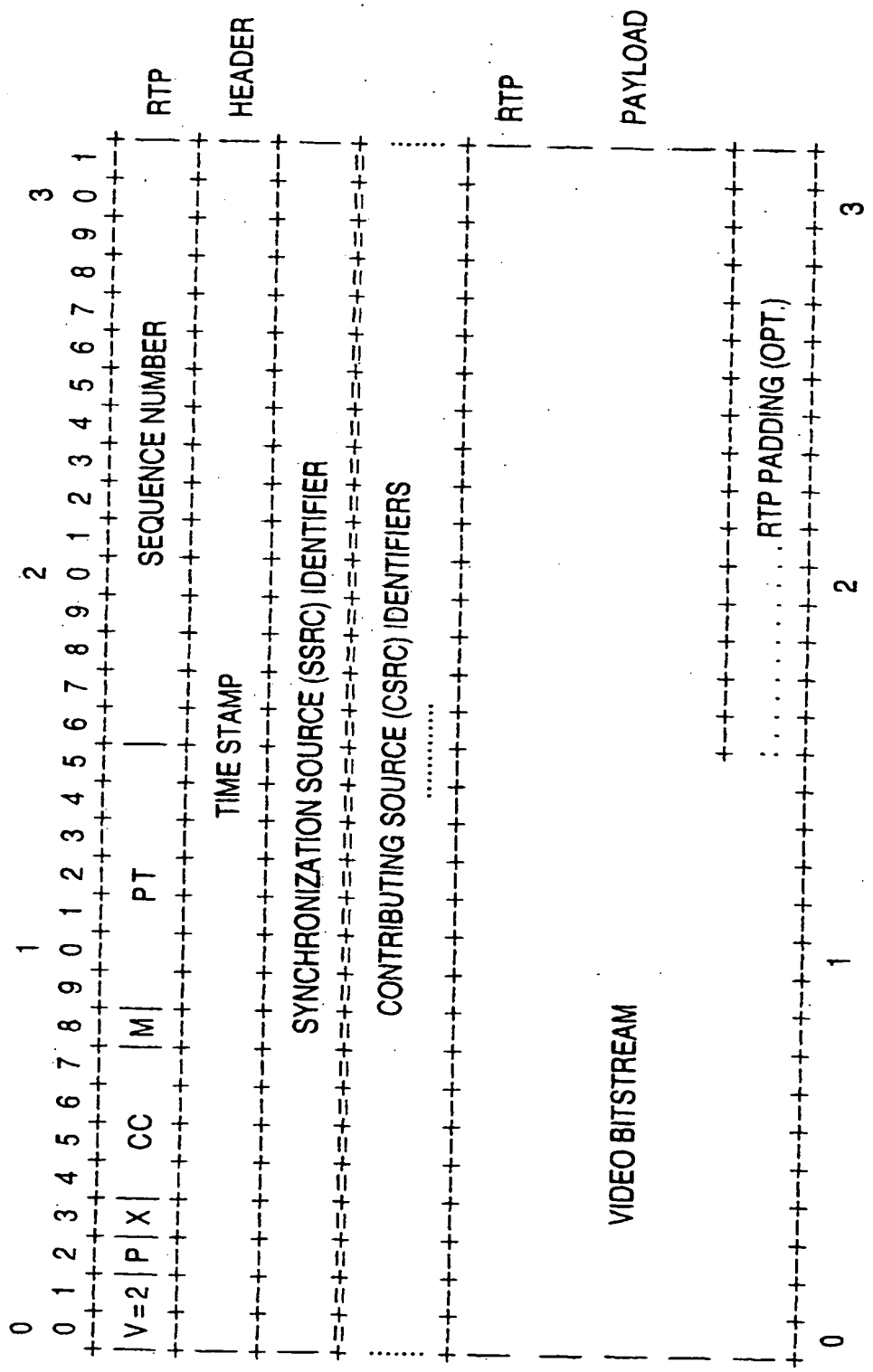


FIG. 9

BNSDOCID: <EP_____1035735A2_I_>

FIG. 10



BNSDOCID: <EP_____1035735A2_I_>

FIG. 11A

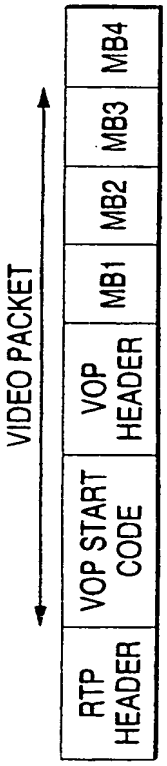


FIG. 11B

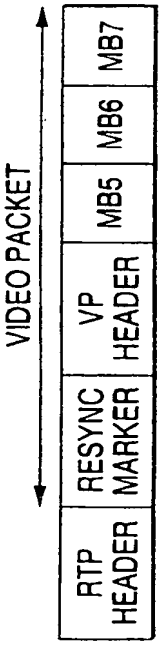


FIG. 11C

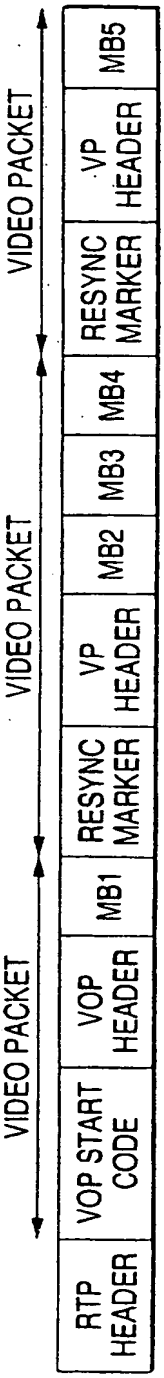


FIG. 11D

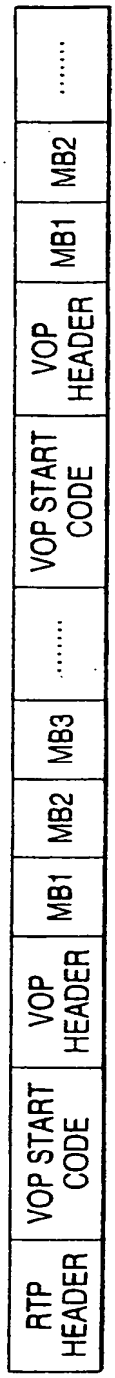


FIG. 11E

BNSDOCID: <EP_____1035735A2_I_>

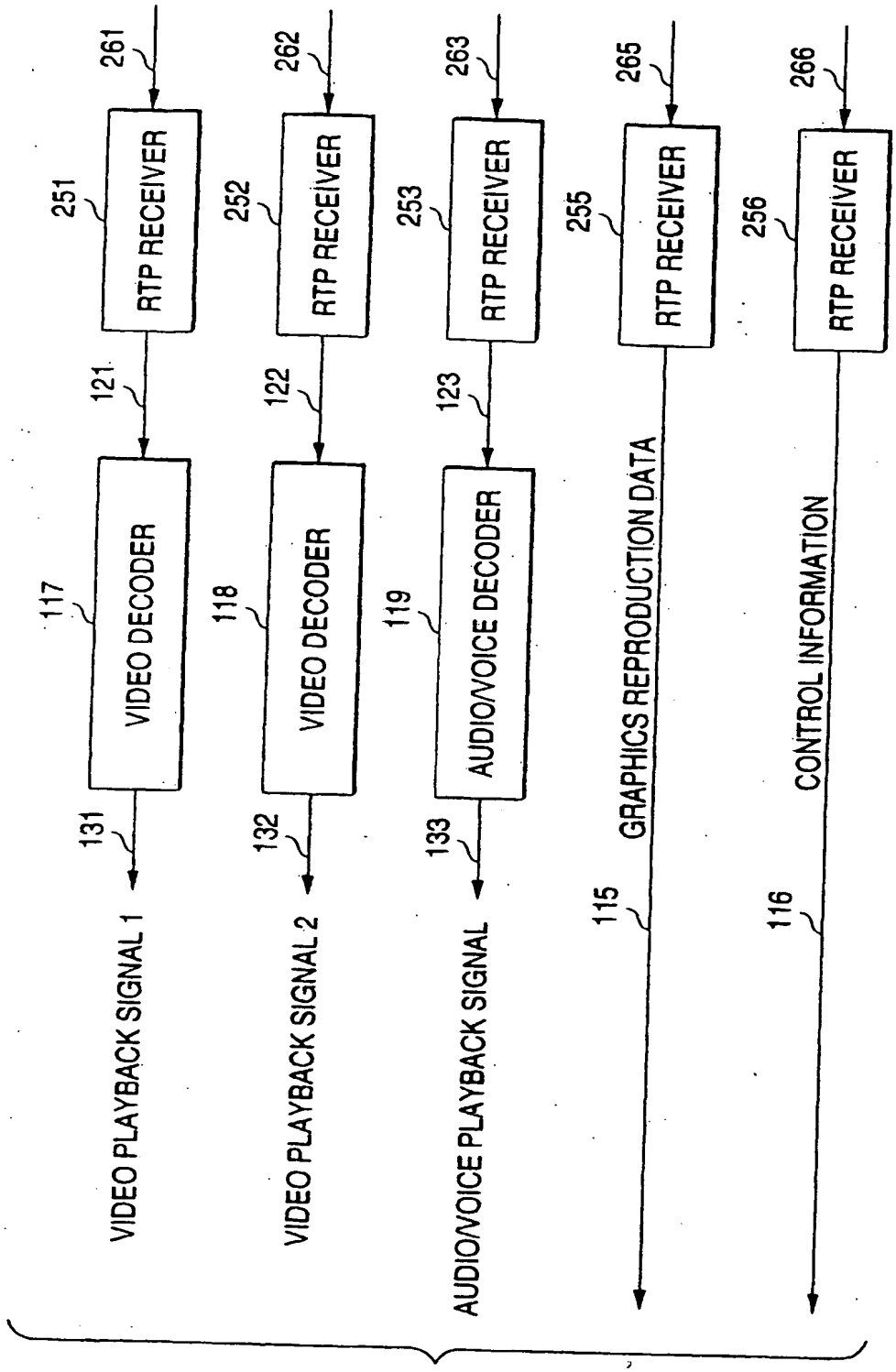
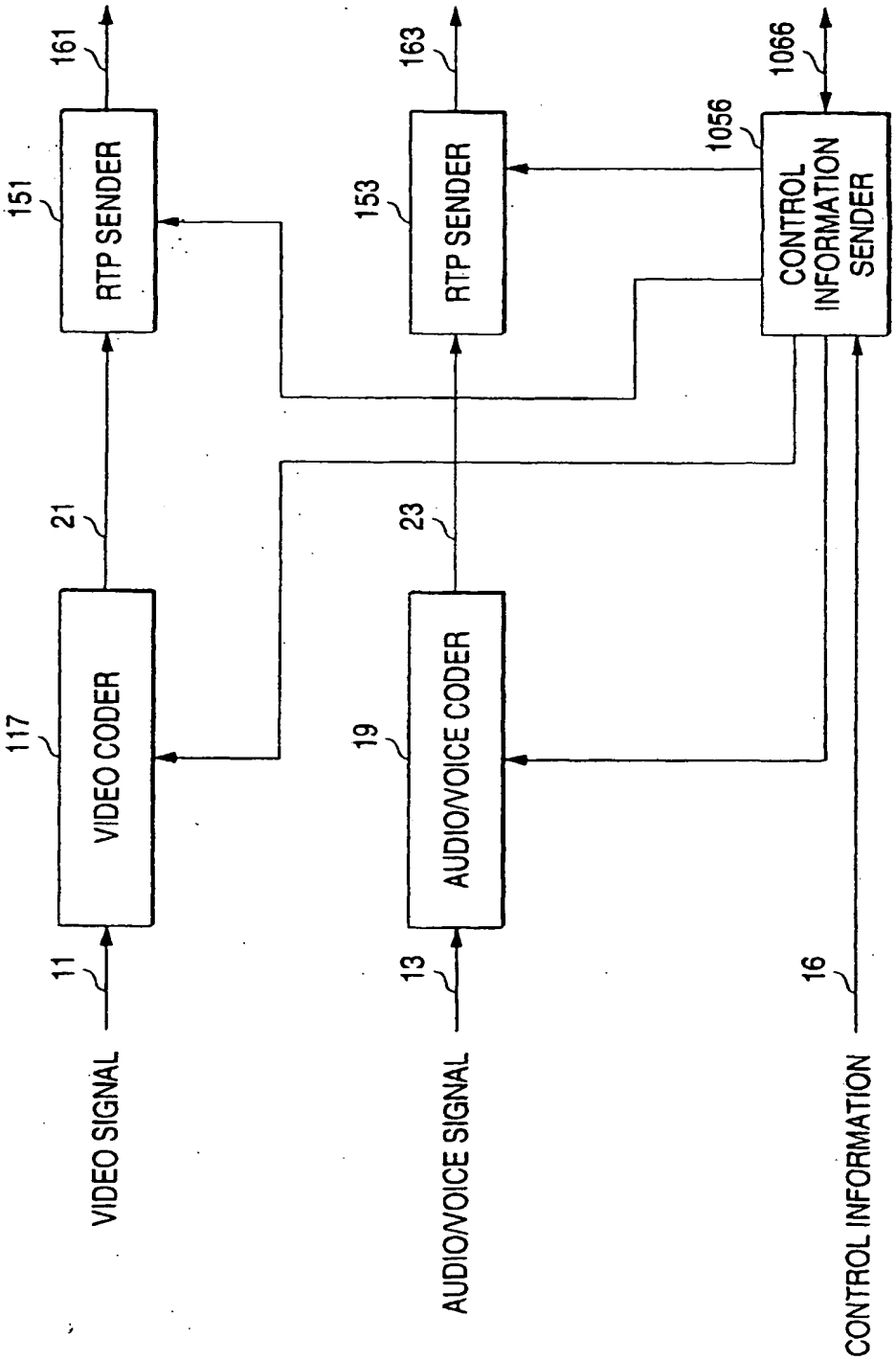


FIG. 12

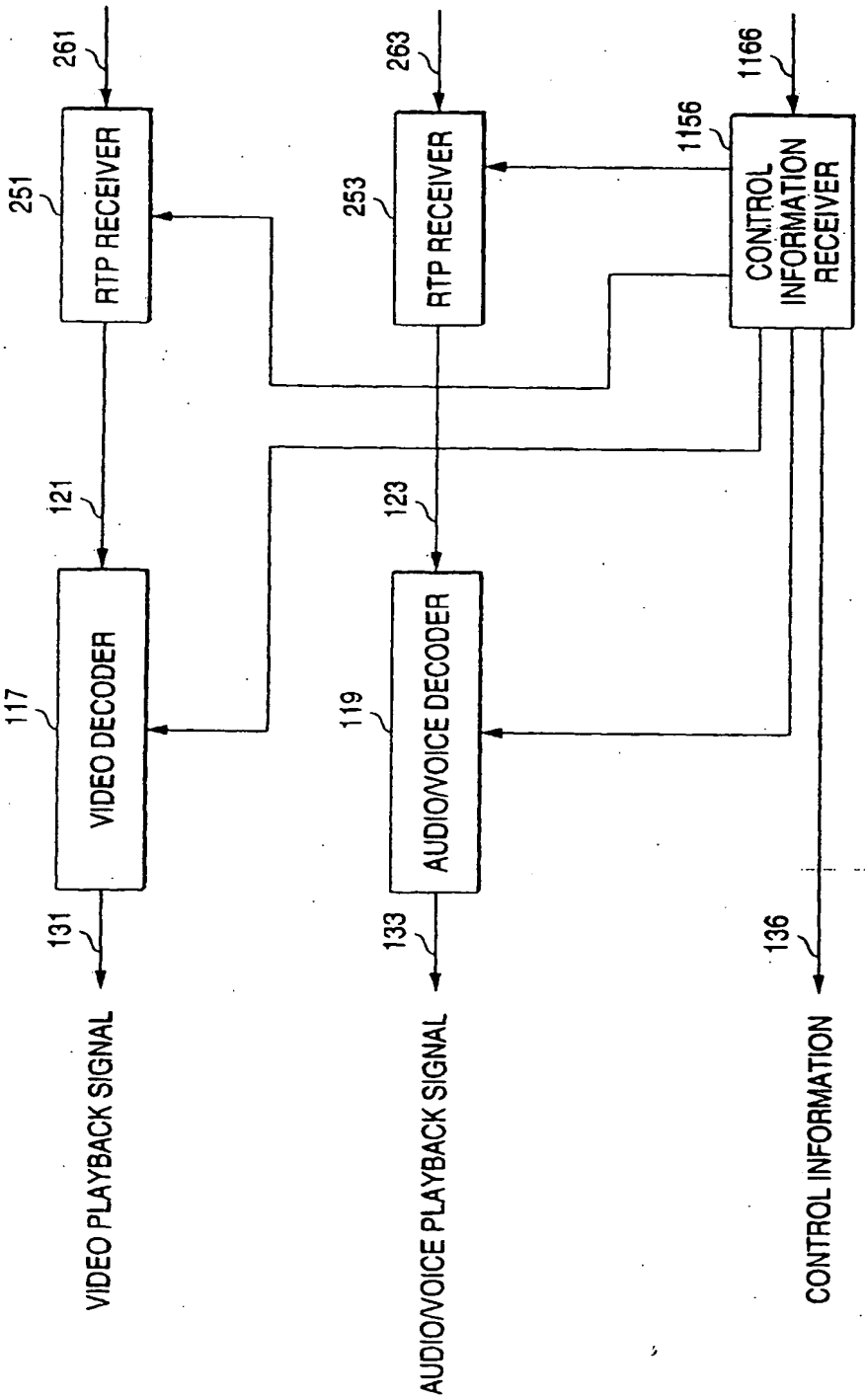
INSDOCID: <EP_____1035735A2_I_>

FIG. 13



BNSDOCID: <EP_____1035735A2_I_>

FIG. 14



BNSDOCID: <EP_____1035735A2_I_>

FIG. 15A

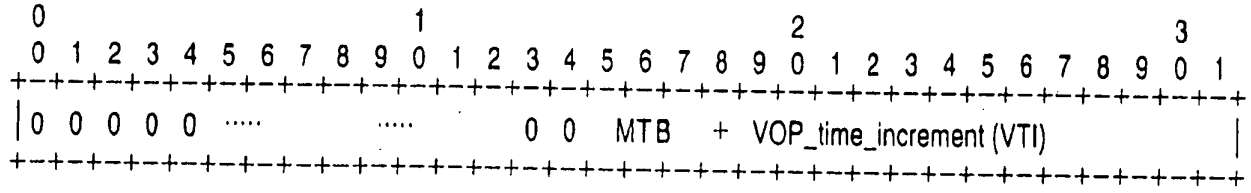


FIG. 15B

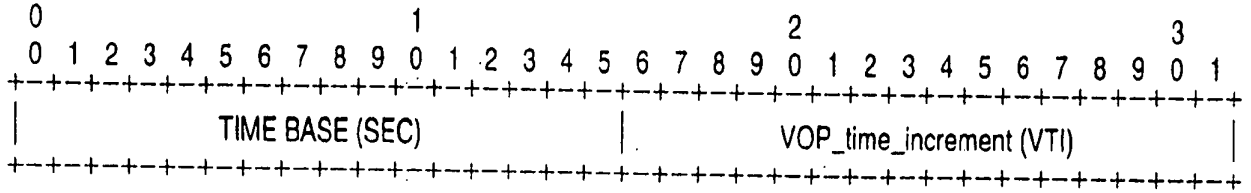


FIG. 15C

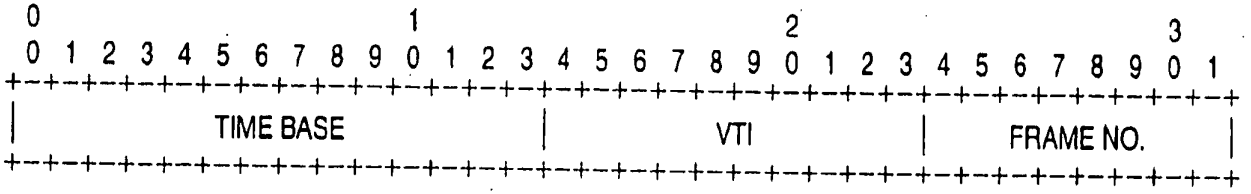


FIG. 15D

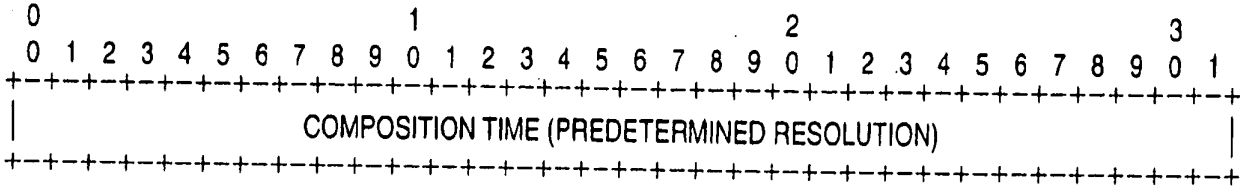


FIG. 15E

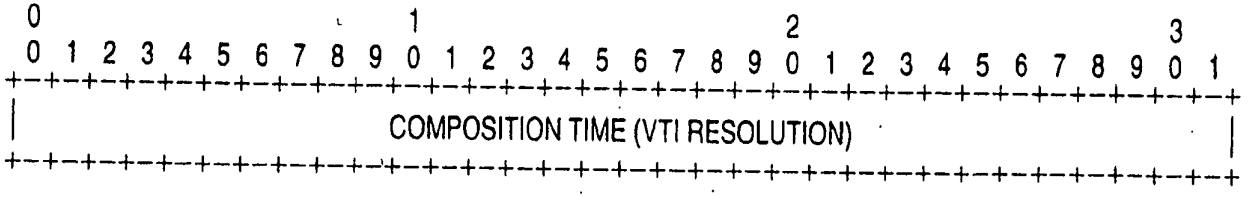
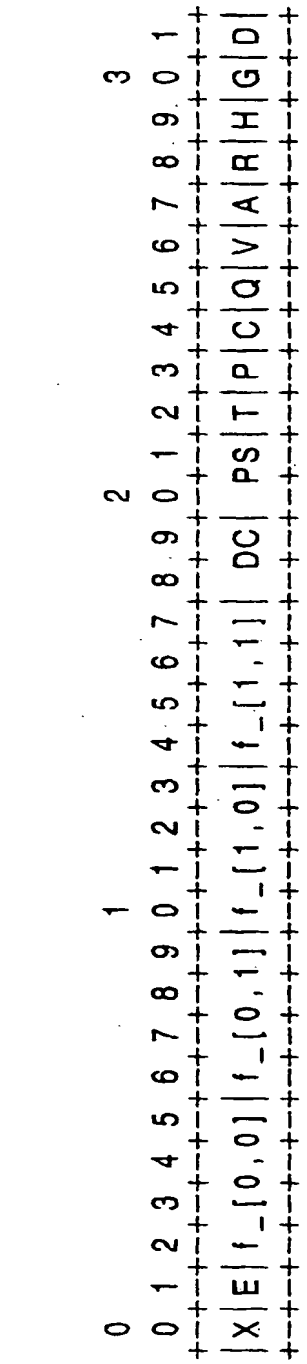


FIG. 16



BNSDOCID: <EP_____1035735A2_I_>

FIG. 17A

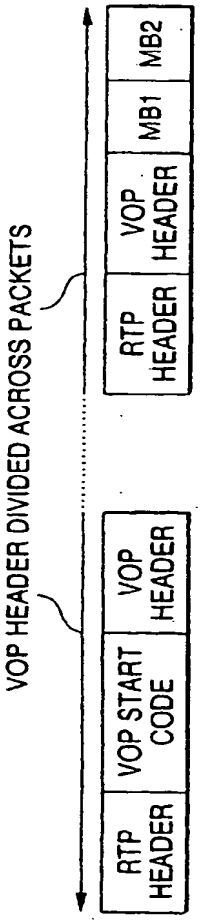


FIG. 17B

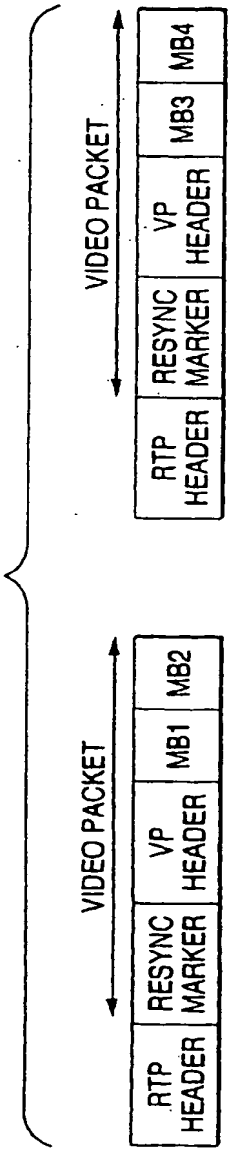
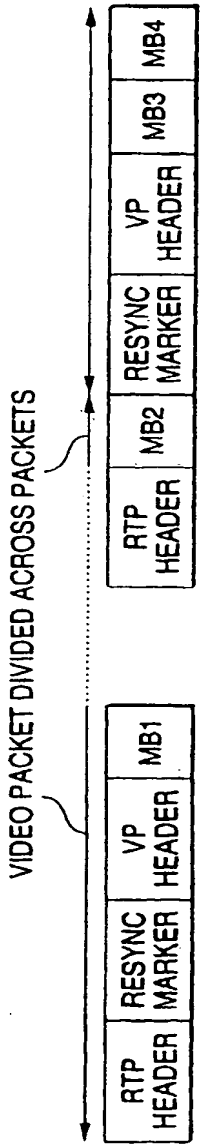
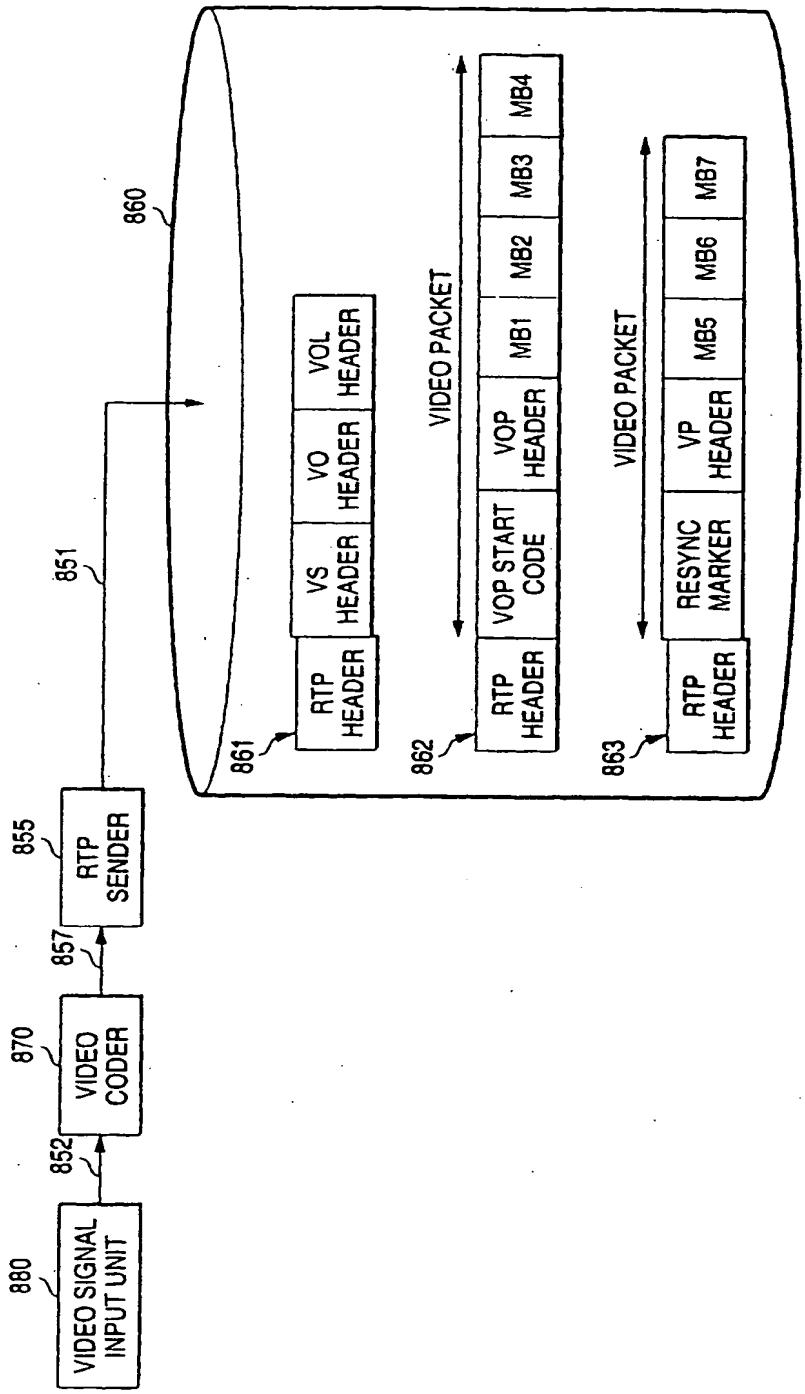


FIG. 17C



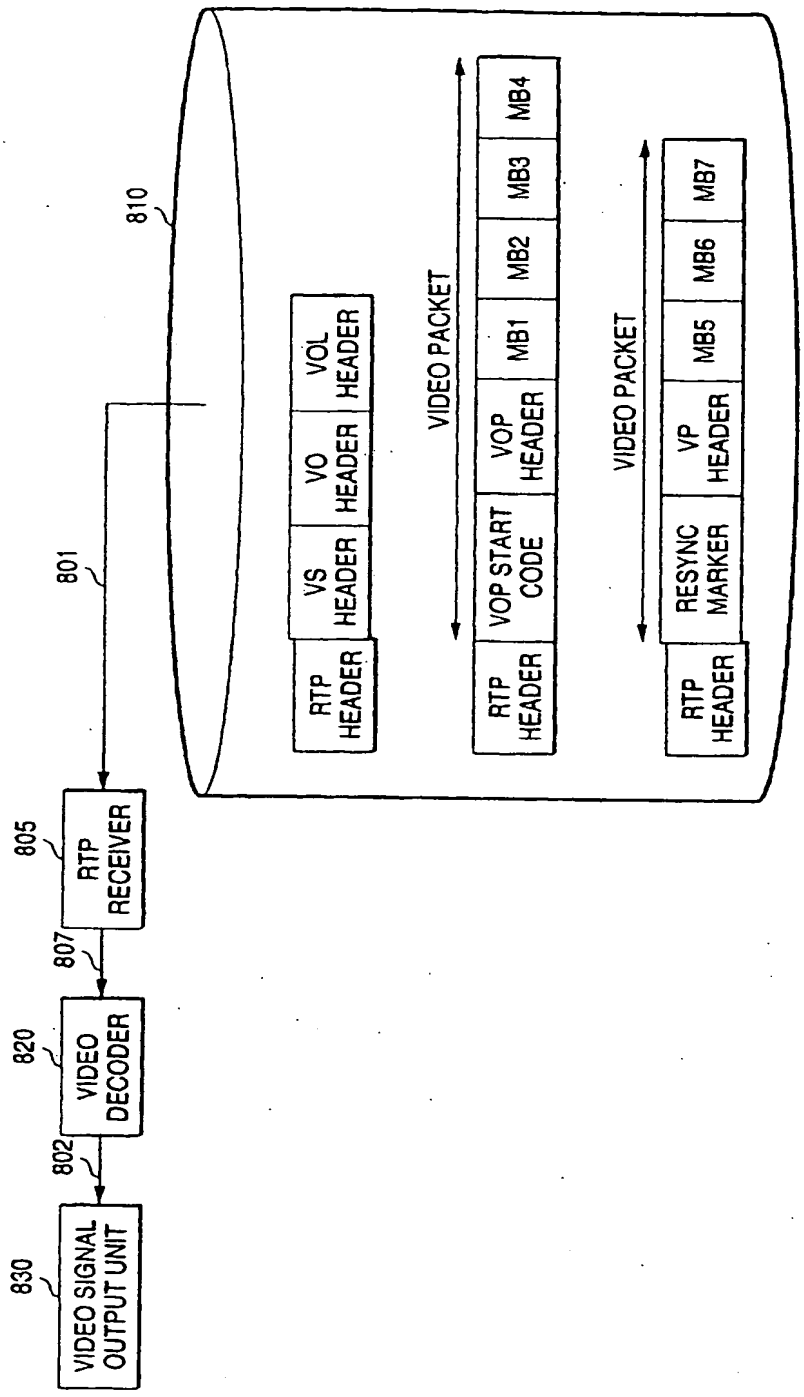
BNSDOCID: <EP_____1035735A2_I_>

FIG. 18



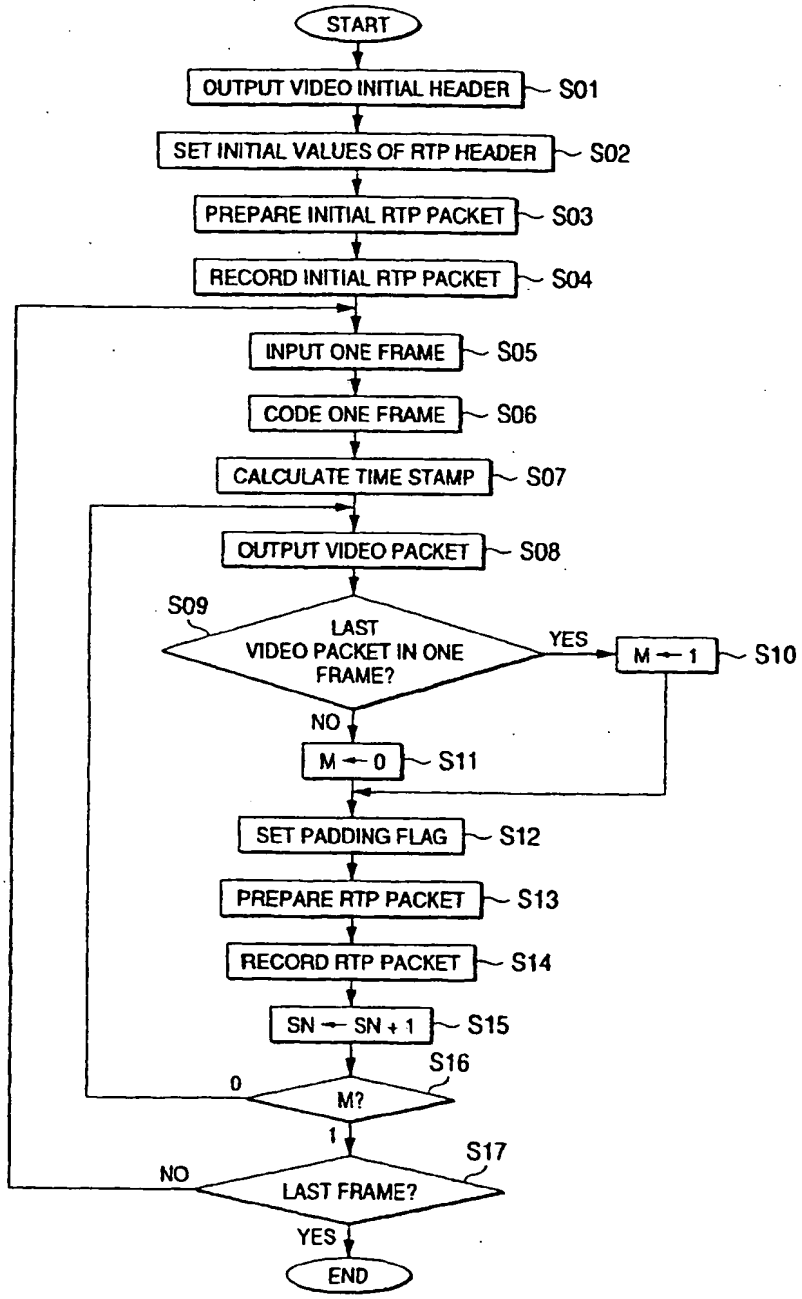
BNSDOCID: <EP_____1035735A2_I_>

FIG. 19



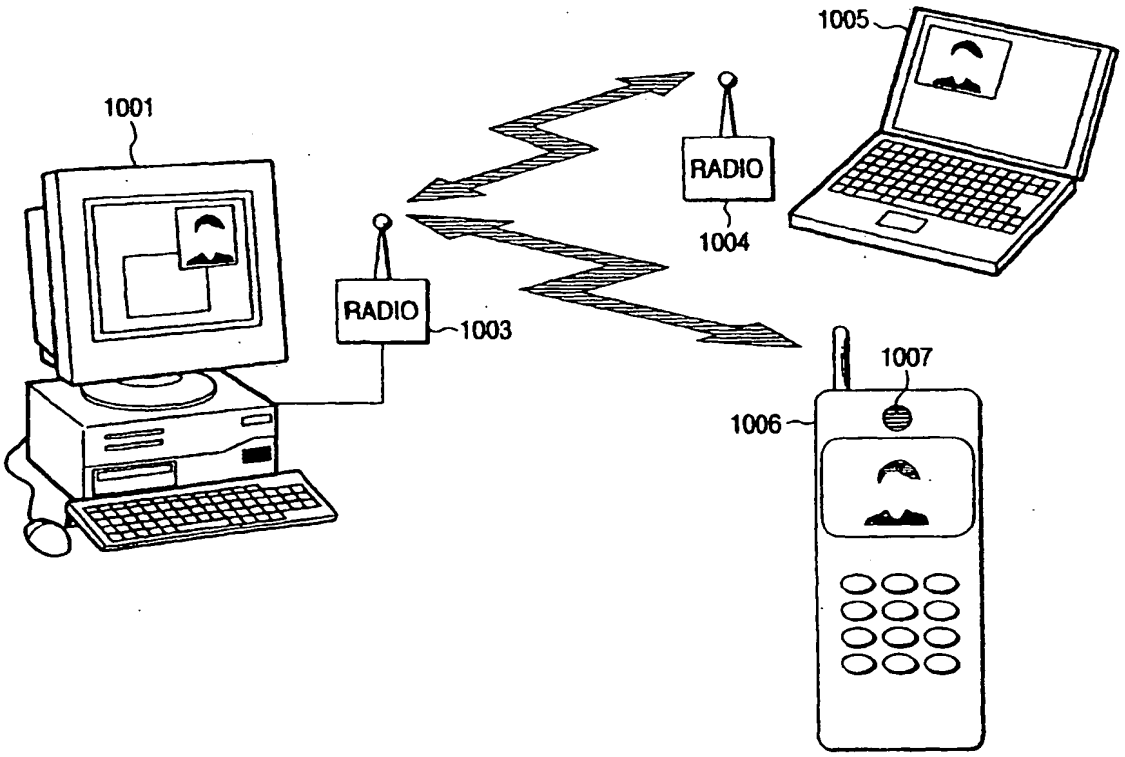
BNSDOCID: <EP_____1035735A2_I_>

FIG. 20



BNSDOCID: <EP_____1035735A2_I_>

FIG. 21



BNSDOCID: <EP_____1035735A2_I_>

(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 0 924 934 A1

(12) EUROPEAN PATENT APPLICATION

(43) Date of publication:
23.06.1999 Bulletin 1999/25

(51) Int. Cl.[6]: H04N 7/52

(21) Application number: 98124300.9

(22) Date of filing: 21.12.1998

(84) Designated Contracting States:
AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU MC NL PT SE
Designated Extension States:
AL LT LV MK RO SI

(30) Priority: 22.12.1997 JP 36479497

(71) Applicant: NEC CORPORATION
Tokyo (JP)

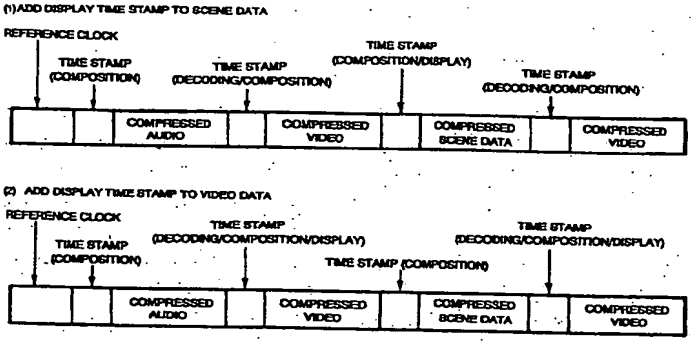
(72) Inventor: Katto, Jiro
Minato-ku, Tokyo (JP)

(74) Representative:
VOSSIUS & PARTNER
Siebertstrasse 4
81675 München (DE)

(54) Coding/decoding apparatus, coding/decoding system and multiplexed bit stream

(57) A coding apparatus of the present invention comprises coding circuit 1 for audio signals, coding circuit 2 for video signals, interface circuit 3 on input of scene data, coding circuit 4 for scene data, composition circuit 5, multiplexing circuit 6, display circuit 7 and clock generating circuit 8. Each of coding circuits 1, 2 and 4 outputs time information representing a decoding timing, and composition circuit 5 outputs time information representing a composition timing. Multiplexing circuit 6 multiplexes time information together with the compressed data given from each of coding circuits 1, 2 and 4, thereby generating a bit stream.

FIG.26



EP 0 924 934 A1

Description

[0001] The present invention relates to a coding/decoding apparatus, a coding/decoding system and a multiplexed bit stream and particularly, to a system for synchronously combining and reproducing natural pictures, voices, and computer graphics.

[0002] MPEG (Motion Picture Coding Expert Group) has been known as an international standard for coding standardization for compressing, multiplexing and transferring or storing audio signal (or voice signal), video signal, and artificial scene data such as computer graphic, and then separating and expanding the signals and data to obtain original signals. The MPEG is defined by the working group (WG) 11 within SC29 which are managed under JTC1 (Joint Technical Committee 1) for handling common items in data processing fields of ISO (International Organization for Standardization) and IEC (International Electrotechnical Commission). In the MPEG, a mechanism for synchronously reproducing each media from multiplexed data is described.

[0003] First, a mechanism for synchronously reproducing an audio signal and a video signal from multiplexed data is described in ISO/IEC 13818-1 "Information Technology Generic Coding of Moving Pictures and Associated Audio Systems" (popularly called MPEG-2 Systems). Fig. 53 of the accompanying drawings shows the construction of a fixed delay model used for the description. This figure shows an abstracted system architecture when MPEG-2 is applied to compress audio signals and video signals.

[0004] In Fig. 53, encoder 71 compresses (encodes) audio signal, and encoder 72 compresses (encodes) video signal. Buffer 73 buffers the audio data compressed by the encoder 71, and buffer 74 buffers the video data thus compressed by the encoder 72. Multiplexing circuit 75 multiplexes the compressed audio data stored in the buffer 73 and compressed video data stored in the buffer 74. At this time, a reference clock that is needed for synchronous reproduction and time stamps are embedded as additive information into the multiplexed data.

[0005] Specifically, the time stamps are a decoding time stamp representing a decoding timing and a display time stamp representing a display timing. The decoding time stamp is generally used only when interpolative prediction is carried out. This is because when the interpolative prediction is carried out, the decoding timing and the display timing are different from each other in some cases. In the other cases, the decoding time stamp is unnecessary.

[0006] Storage/transmission device 76 stores or transmits the multiplexed data created by the multiplexing circuit 75. Separation circuit (demultiplexing circuit) 77 separates compressed audio data, compressed video data, and a reference clock and time stamp used for synchronous reproduction from the multiplexed data supplied from the storage/transmission device 76. Buffer 78 buffers the compressed audio data supplied from the separation circuit 77, and buffer 79 buffers the compressed video data supplied from the separation circuit 77. Decoder 80 decodes and reproduces the compressed audio data stored in the buffer 78, and decoder 81 decodes and displays the compressed video data stored in the buffer 79.

[0007] The synchronous reproduction of the audio signals and video signals in Fig. 53 is implemented as follows. The reference clock embedded in the multiplexed data is used to control the oscillation frequency of a clock generating circuit for driving the decoder 80 and decoder 81, and PLL (Phased Locked Loop) is generally used. The synchronization between the encoder side and the decoder side is established by the PLL. The time stamp embedded in the multiplexed data is used to transmit the decoding timing of the decoder 80 and decoder 81 or the reproduction/display timing of the decoding result. The time axes of the encoder side and decoder side are synchronized with each other with a fixed delay being set therebetween by the reference clock, and the decoding operation is started at the time which is intended at the encoder side and the reproduction/display is carried out.

[0008] Accordingly, the synchronous reproduction of the audio signals and video signals can be implemented insofar as a suitable time stamp is set at the encoder side. In the case of an application in which synchronous reproduction isn't needed between the encoder side and the decoder side, the synchronous reproduction is carried out with the clock of the decoder itself without using the reference clock.

[0009] Next, ISO/IEC JTC1/SC29/WG11 N1825 "Working Draft 5.0 of ISO/IEC 14996-1" (popularly called MPEG-4 Systems) describes a mechanism for synchronously reproducing audio signals, video signals, and artificial scene data such as computer graphics from multiplexed data.

[0010] Fig. 54 shows a system decoder model (SDM) used for the description of the above mechanism. This model is an abstracted system decoder when MPEG-4 is applied to compress audio signals, video signals, and artificial scene data such as computer graphics. In this paper, detailed description isn't made on the model and concrete construction of the encoder, however, it is described as syntax that a reference clock and a time stamp are embedded as additive information in multiplexed data. Specifically, there are provided two time stamps, a decoding time stamp representing a decoding timing and a composite time stamp representing a timing at which decoding data can be supplied to a composition circuit.

[0011] In Fig. 54, a separation circuit 91 separates from the multiplexed data compressed audio data, compressed video data, compressed scene data, and a reference clock and a time stamp used for synchronous reproduction. Buffer 92 buffers the compressed audio

BNSDOCID: <EP_____0924934A1_I_>

data supplied from the separation circuit 91, and buffer 93 buffers the compressed video data supplied from the separation circuit 91. Buffer 94 buffers the compressed artificial scene data supplied from the separation circuit 91. Decoder 95 decodes the compressed audio data stored in the buffer 92, decoder 96 decodes the compressed video data stored in the buffer 93, and decoder 97 decodes the compressed artificial scene data stored in the buffer 94.

[0012] Buffer 98 buffers the audio signal decoded by the decoder 95, buffer 99 buffers the video signal decoded by the decoder 96, and buffer 100 buffers the artificial scene data decoded by the decoder 97. Composition circuit 101 composes a scene on the basis of the audio signal stored in the buffer 98, the video signal stored in the buffer 99 and the artificial scene data stored in the buffer 100. At this time, the scene information that is composed is described in the artificial scene data, and in accordance with the scene information the audio signal is modulated or the video signal is deformed, and the signal is mapped to an object in the scene. Display circuit 102 reproduces/displays a scene supplied from the composition circuit 101.

[0013] The composite and reproduction of the audio signal, the video signal and the artificial scene data in Fig. 54 is implemented as follows:

[0014] The reference clock can be provided every decoder. After it is picked up from the multiplexed data, it is input to a clock generating circuit which is provided every decoder in order to control the oscillation frequency of the clock generating circuit, whereby the synchronization between the encoder side and the decoder side can be established every decoder. The time stamp can be also provided every decoder. After it is picked up from the multiplexed data, it is used to transmit the time at which the decoding timing of the decoder or the decoding result can be supplied to the composition circuit 101. The time axes of the encoder side and the decoder side are synchronized with each other with a fixed delay being set therebetween by the reference clock, and the decoding is started at the time intended by the encoder side and the writing operation into the buffer is carried out.

[0015] Subsequently, the composition circuit 101 takes out the audio signal, the video signal and the artificial scene data held in each buffer to perform scene composition. The times at which the audio signal, the video signal and the scene data are obtained by the composition circuit 101 are respectively given on the basis of the composite time stamps added to these signals and data. However, the timing for composing a scene is unclear, and the composition circuit 101 itself is set to start a event processing in accordance with a discrete time event described in the scene data. Finally, the display circuit 102 reproduces and displays the scene supplied from the composition circuit 101.

[0016] Further, as representative one of artificial scene data, VRML (Virtual Reality Modeling Language) has been known as a description format to describe computer graphics, transmit or store the data thus described, build and share a virtual three-dimensional space on the of the data. VRML is defined as international standards by SC24 managed under JTC1 (Joint Technical Committee 1) for handling common items in the data processing fields of ISO (International Organization for Standardization) and IEC (International Electrotechnical Commission) and a VRML consortium to which associated companies pertain in cooperation with each other. In this VRML, a description method of taking an audio signal and a video signal into a scene is further described.

[0017] The details of the description method are described in ISO/IEC DIS 14772-1 "The virtual Reality Modeling Language (popularly called VRML97). IN the ISO/IEC DIS 14772-1, not only computer graphics, but also ISO/IEC 11172 (popularly called MPEG-1) which is one of the MPEG standards are contained as support targets. MPEG-1 is one of coding international standards for audio signals and video signals. Specifically, the audio signals and the video signals are mapped as a sound source and as a moving picture texture for a three-dimensional object respectively in a three-dimensional scene constructed by VRML. Further, the description of a time event is supported on VRML, and a time event occurs according to a time stamp described in the VRML format.

[0018] The time event is further classified into two types of a continuous time event and a discrete time event. The continuous time event is an event in which the action of an animation or the like is continuous on time axis, and the discrete time event is an event in which an object in a scene starts after a time elapses.

[0019] Fig. 55 shows the construction of a decoding processing system for receiving the VRML format and constructs a three-dimensional scene (called as "Browser" in VRML). Buffer 111 receives through the internet multiplexed data compressed by MPEG-1 and buffers the data received. Buffer 112 receives through the internet the VRML format or the compressed VRML format and buffers the format received. At this time, the original place of the VRML format may be different from that of the MPEG-1 data.

[0020] Separation circuit 113 separates compressed audio data and compressed video data from the MPEG-1 multiplexed data supplied from the buffer 111. Decoder 114 decodes the compressed audio data supplied from the separation circuit 113, and decoder 115 decodes the compressed video data supplied from the separation circuit 114. Decoder 116 decodes the compressed VRML format stored in the buffer 112. When the VRML format is not compressed, no action is taken. Memory 117 stores the audio signal decoded by the decoder 114, and memory 118 stores the video signal decoded by the decoder 115. Memory 119 stores the VRML format decoded by the decoder 116.

[0021] Composition circuit 120 synthesizes a scene

BNSDOCID: <EP_____0924934A1_I_>

on the basis of the audio signal stored in the memory 117, the video signal stored in the memory 118 and the artificial scene data stored in the memory 119. In this case, scene information to be composed is described in the artificial scene data. According to the scene information, the audio signal is modulated and the video signal is deformed, and then these signals are mapped into an object in the scene. Display circuit 121 reproduces/displays the scene supplied from the composition circuit 120.

[0022] The composite of the audio signal, the video signal and the VRML format in Fig. 55 and the reproduction thereof are implemented as follows:

[0023] After the loading of the MPEG-1 multiplexed data from the external to the buffer 111 is terminated, the decoder 114 decodes the compressed audio data and the decoder 115 decodes the compressed video data, and the audio signal and the video signal obtained through the above decoding operation are written into the memory 117 and the memory 118 respectively. Further, after the loading of the VRML format from the external to the buffer 112 is terminated, the decoder 116 decodes the VRML format when the VRML format is compressed or takes no action when the VRML format is not compressed, and then writes the VRML format thus obtained into the memory 119. After the above processing is terminated, that is, the processing of a part surrounded by a dotted line indicated by reference numeral 222 is terminated, the composition circuit 120 and the display circuit 121 start operating to perform composite (mixing), reproduction and display.

[0024] On the other hand, when it is intended that only the video signal and the computer graphics are combined with each other, a chromakey system which has been already used for the weather forecast in the present broadcasting system has been known. According to the chromakey system, a person or an object is disposed under the background whose color is specified to a single color such as blue color or the like to shoot an overall picture, and then the background-colored portion is deleted from the picture, whereby only the person or the object in front of the background can be picked up.

[0025] Fig. 56 shows the construction of a coding processing system for creating a composite picture of the video signal and the computer graphics by using the chromakey system, and compressing and multiplexing the composite picture and the audio signal. Chromakey processing circuit 131 deletes from an input video signal a portion having the color coincident with the background color. Composition circuit 132 creates a computer graphics image from artificial scene data given. Memory 133 stores a cut-out picture supplied from the chromakey processing circuit 131. In this case, memory 133 may store directly the picture data and inform merely a subsequent-stage convolution circuit 135 that the RGB value corresponding to the background color is deleted. Memory 134 stores the computer graphics picture generated by the composition circuit 132. The convolution circuit 135 overwrites the cut-out picture obtained from the memory 133 on the computer graphics image obtained from the memory 134. It may be also allowed to detect the RGB value corresponding to the background color and replace only pixels located within a specified range by a computer graphics image.

[0026] Encoder 136 compresses (encodes) the audio signal. Encoder 137 compresses the composite picture obtained from the convolution circuit 135. Buffer 138 buffers the audio data compressed by the encoder 136, and buffer 139 buffers the composite picture data compressed by the encoder 137. Multiplexing circuit 140 multiplexes the compressed audio data stored in the buffer 138 and the compressed composite picture data stored in the buffer 139. At this time, the reference clock which is necessary for the synchronous reproduction and the time stamp are embedded as additive information into the multiplexed data.

[0027] The creation of the composite picture of the video signal and computer graphics is performed in the portion surrounded by a dotted line indicated by reference numeral 141. The other portions correspond to the coding portion of the coding/decoding system shown in Fig.53. That is, the video signal and the computer graphics are first combined with each other to obtain a composite picture, and then the composite picture and the audio signal are compressed and multiplexed. The construction of the decoding side is the same as that of Fig. 53.

[0028] The coding/decoding synchronous reproduction system of the audio signal and the video signal shown in Fig. 53 relates to the coding, multiplexing, separating and decoding for the audio signal and the video signal, and no description is made on the processing of artificial scene data such as computer graphics.

[0029] Further, in the decoding synchronous reproduction system of the audio signal, the video signal and the artificial scene data shown in Fig. 54, the decoding timing and the timing at which each data may be supplied to the composition circuit are given. However, the timing at which all the data are composed and the timing at which the composite picture is displayed are not specified. In other words, the composition circuit is set to start its composite operation freely. Further, it is suggested that the composition (mixing) is started in accordance with a discrete time event described in the artificial scene data.

[0030] However, the artificial scene data suffers a buffer delay in the decoding operation, and thus a desired time may have passed at the time when the artificial scene data are supplied to the composition circuit 101. Therefore, the artificial scene data itself cannot be used to give an accurate timing for composing. Further, when a continuous time event is described in the artificial scene data, the composition start time is different between the coding side and the decoding side in some cases. Therefore, occurrence of an accurately coinci-

3NSDOCID: <EP_____0924934A1_I_>

dent continuous time event cannot be ensured. Particularly, in the case of animation or the like for which motion is required to be continuously represented, the position of a moving object is displaced between the coding side and the decoding side. Due to the above problem, a composite picture desired by the coding side cannot be composed while it is accurately coincident at the decoding side.

[0031] Further, the decoding and reproducing system of the audio signal, the video signal and the artificial scene data shown in Fig. 55 does not support stream data which are transmitted continuously on time axis. That is, the processing of a portion 122 surrounded by a dotted line must be finished before the reproduction is started.

[0032] Still further, in the coding/decoding synchronous reproducing system of the audio signal, the video signal and the artificial scene data shown in Fig. 56, the composite picture is degenerated into a mere two-dimensional picture at the coding side, and thus an interaction function which would be obtained by using the artificial scene data is lost. That is, there is a disadvantage that additive functions such as movement of a visual point in the three-dimensional space, and navigation cannot be implemented.

[0033] An object of the present invention is to provide a coding apparatus, a decoding apparatus, a coding/decoding system and a multiplexed bit stream which implements coding/decoding synchronous reproduction of an audio signal, a video signal and artificial scene data while excluding the disadvantage of the conventional systems described above, ensuring generation of a composite picture desired at the coding side, supporting stream data transmitted continuously on time axis, and supporting the interaction function in the decoding side.

[0034] A coding apparatus according to the present invention comprises: audio signal coding means for coding an audio signal; video signal coding means for coding a video signal; interface means for accepting information on a composite scene; scene data coding means for coding scene data supplied from the interface means; composition means for composing a scene from the audio signal supplied from the audio signal coding means, the video signal supplied from the video signal coding means and the composite scene data supplied from the scene data coding means; display means for reproducing/displaying the composite picture signal and the audio signal supplied from the composition means; clock supply means for supplying clocks to the audio signal coding means, the video signal coding means, the scene data coding means and the composition means; and multiplexing means for creating a bit stream on the basis of the time information and compressed audio data supplied from the audio signal coding means, the time information and compressed video data supplied from the video signal coding means, the time information and compressed scene data supplied from the scene data coding means, the time information supplied from the composition means and the clock value supplied from the clock supplying means.

[0035] According to the present invention, the coding apparatus further comprises means for detecting the status of the composition means and controlling the operation of the coding means of the video signal.

[0036] According to the present invention, the coding apparatus further comprises means for detecting the status of the coding means for the audio signal, the status of the coding means for the video signal and the status of the coding means for the scene data, and controlling the operation of the composition means.

[0037] According to the coding apparatus of the present invention, the clock supply means includes first clock supply means for supplying clocks to the audio signal coding means, second clock supply means for supplying clocks to the video signal coding means and third clock supply means for supplying clocks to the scene data coding means and composition means, and the multiplexing means multiplexes the clock values supplied from the first, second, and third clock supply means respectively.

[0038] According to the coding apparatus of the present invention, the clock supply means includes first clock supply means for supplying clocks to the audio signal coding means, second clock supply means for supplying clocks to the video signal coding means and composition means, and third clock supply means for supplying clocks to the scene data coding means, and the multiplexing means multiplexes the clock values supplied from the first, second, and third clock supply means respectively.

[0039] A decoding apparatus according to the present invention comprises: means for separating both of compressed data and time information of an audio signal, both of compressed data and time information of a video signal, both of compressed data and time information of scene data, time information of scene composition and clock information from a bit stream; means for decoding the audio signal on the basis of the compressed data and time information of the audio signal; means for decoding the video signal on the basis of the compressed data and time information of the video signal; means for decoding the scene data on the basis of the compressed data and time information of the scene data; means for composing a scene on the basis of the time information for the scene composition supplied from the separation means, the audio signal supplied from the decoding means for the audio signal, the video signal supplied from the decoding means for the video signal and the scene data supplied from the decoding means for the scene data; means for generating clocks according to the clock value supplied from the separating means and supplying the clocks to the decoding means for the audio signal, the decoding means for the video signal, the decoding means for the scene data and the composition means; means for reproducing/dis-

BNSDOCID: <EP_____0924934A1_I_>

playing the composite picture signal and the audio signal supplied from the composition means; and interface means for accepting an interaction from a viewer to the composite picture.

[0040] According to a first embodiment of the decoding apparatus, the separation means separates a plurality of independent clock values from the bit stream, and the independent clock values are input to means for supplying the clocks to the decoding means for the audio signal, means for supplying the clocks to the decoding means for the video signal, and means for supplying the clocks to the decoding means for the scene data and the composition means.

[0041] According to a second embodiment of the decoding apparatus, the separation means separates a plurality of independent clock values from the bit stream, and the independent clock values are input to means for supplying the clocks to the decoding means for the audio signal, means for supplying the clocks to the decoding means for the video signal and the composition means, and means for supplying the clocks to the decoding means for the scene data.

[0042] A multiplexed bit stream according to the present invention comprises an audio signal, a video signal and scene data, characterized in that a flag representing whether time information representing a decoding timing doubles as time information representing a composition timing is added to said time information.

Fig. 1 is a block diagram showing a first embodiment of a coding apparatus according to the present invention;

Fig. 2 is a block diagram showing the construction of a coding circuit of Fig. 1;

Fig. 3 is a first block diagram showing the construction of a composition circuit of Fig. 1;

Fig. 4 is a block diagram showing the construction of a multiplexed circuit of Fig. 1;

Fig. 5 is a block diagram showing a second embodiment of the coding apparatus according to the present invention;

Fig. 6 is a block diagram showing the construction of a coding circuit of Fig. 5;

Fig. 7 is a first block diagram showing the construction of a composition circuit of Fig. 5;

Fig. 8 is a block diagram showing a third embodiment of the coding apparatus according to the present invention;

Fig. 9 is a block diagram showing the construction of a coding circuit of Fig. 8;

Fig. 10 is a first block diagram showing the construction of a composition circuit of Fig. 8;

Fig. 11 is a block diagram showing a fourth embodiment of the coding apparatus according to the present invention;

Fig. 12 is a block diagram showing the construction of a multiplexed circuit of Fig. 11;

Fig. 13 is a block diagram showing a fifth embodiment of the coding apparatus according to the present invention;

Fig. 14 is a block diagram showing a sixth embodiment of the coding apparatus according to the present invention;

Fig. 15 is a block diagram showing a seventh embodiment of the coding apparatus according to the present invention;

Fig. 16 is a block diagram showing an eighth embodiment of the coding apparatus according to the present invention;

Fig. 17 is a block diagram showing a ninth embodiment of the coding apparatus according to the present invention;

Fig. 18 is a block diagram showing a first embodiment of a decoding apparatus according to the present invention;

Fig. 19 is a block diagram showing the construction of a separation circuit of Fig. 18;

Fig. 20 is a block diagram showing the construction of a decoding circuit of Fig. 18;

Fig. 21 is a first block diagram showing the construction of a composition circuit of Fig. 18;

Fig. 22 is a block diagram showing a second embodiment of the decoding apparatus according to the present invention;

Fig. 23 is a block diagram showing the construction of a separation circuit of Fig. 22;

Fig. 24 is a block diagram showing a third embodiment of the decoding apparatus according to the present invention;

Fig. 25 is a block diagram showing a coding/decoding system according to the present invention;

Fig. 26 is a diagram showing a bit stream generated by the coding apparatus according to the first embodiment of the present invention;

Fig. 27 is a diagram showing a bit stream generated by the coding apparatus according to the fourth embodiment of the present invention;

Fig. 28 is a time chart for normal coding, decoding and composition;

Fig. 29 is a time chart for coding, decoding and composition when excessive time is needed for composition;

Fig. 30 is a time chart for coding, decoding and composition, which is solved by the coding apparatus of the second embodiment of the present invention;

Fig. 31 is a time chart for normal coding, decoding and composition in the case of plural inputs;

Fig. 32 is a first time chart for coding, decoding and composition when excessive time is needed for composition in the case of plural inputs;

Fig. 33 is a first time chart for coding, decoding and composition in the case of plural inputs, which is solved by the coding apparatus of the second embodiment of the present invention;

BNSDOCID: <EP_____0924934A1_I_>

Fig. 34 is a second time chart for coding, decoding and composition when excessive time is needed for composition in the case of plural inputs;

Fig. 35 is a second time chart for coding, decoding and composition in the case of plural inputs, which is solved by the coding apparatus of the second embodiment of the present invention;

Fig. 36 is a time chart for coding, decoding and composition, which is solved by the coding apparatus of the third embodiment of the present invention;

Fig. 37 is a time chart for coding, decoding and composition in the case of plural inputs, which is solved by the coding apparatus of the third embodiment of the present invention;

Fig. 38 is a diagram showing data flow among a buffer in a decoding circuit, a memory in the decoding circuit and a composition circuit;

Fig. 39 is a time chart for normal decoding and composition;

Fig. 40 is a time chart for decoding and composition when excessive time is needed for composition;

Fig. 41 is a time chart for decoding and composition, which is solved by the decoding apparatus of the first embodiment of the present invention;

Fig. 42 is a time chart for normal decoding and composition in the case of plural inputs;

Fig. 43 is a time chart for decoding and composition when excessive time is needed for composition in the case of plural inputs;

Fig. 44 is a time chart for decoding and composition in the case of plural inputs, which is solved by the decoding apparatus of the first embodiment of the present invention;

Fig. 45 is a second block showing the construction of the composition circuit of Fig. 1;

Fig. 46 is a second block diagram showing the construction of the composition circuit of Fig. 5;

Fig. 47 is a second block diagram showing the construction of the composition circuit of Fig. 8;

Fig. 48 is a second block diagram showing the construction of the composition circuit of Fig. 18;

Fig. 49 is a diagram showing another example of a bit stream generated by the coding apparatus of the first embodiment of the present invention;

Fig. 50 is a diagram showing another example of a bit stream generated by the coding apparatus of the fourth embodiment of the present invention;

Fig. 51 is a block diagram showing a tenth embodiment of the coding apparatus of the present invention;

Fig. 52 is a block diagram showing the fourth embodiment of the decoding apparatus of the present invention;

Fig. 53 is a diagram showing a conventional coding/decoding synchronous reproducing system for audio signals and video signals;

Fig. 54 is a conventional decoding synchronous reproducing system for audio signals, video signals and artificial scene data;

Fig. 55 is a diagram showing a conventional decoding reproducing system for audio signals, video signals and artificial scene data; and

Fig. 56 is a diagram showing a conventional coding/decoding synchronous reproducing system for audio signals, video signals and artificial scene data.

[0043] Preferred embodiments according to the present invention will be described hereunder with respect to the accompanying drawings.

[0044] Fig. 1 is a block diagram showing a first embodiment of a coding apparatus according to the present invention. The coding apparatus shown in Fig. 1 comprises a coding circuit 1 for audio signals (hereinafter referred to as "audio coding circuit"), a coding circuit 2 for video signals (hereinafter referred to as "video coding circuit"), an interface circuit 3 for input of scene data, a coding circuit 4 for scene data (hereinafter referred to as "scene coding circuit"), a composition circuit 5, a multiplexing circuit 6, a display circuit 7 and a clock generating circuit 8.

[0045] The audio coding circuit 1 compresses an audio signal input thereto, and outputs the compressed data, a time stamp representing a decoding timing and audio data which is locally decoded. The video coding circuit 2 compresses a video signal input thereto, and outputs the compressed data, a timestamp representing a decoding timing and video data which are locally decoded. In place of the video signal, text data, graphics data or the like may be coded in some cases.

[0046] The interface circuit 3 for the input of the scene data accepts description, update on composite scenes from a transmitter, and outputs it as scene data. A keyboard input, a mouse input or the like may be used as the interface. The scene coding circuit 4 receives the scene data from the interface circuit 3, and outputs the compressed data of the scene data, a timestamp representing a decoding timing and scene data which are locally decoded. The time stamp generated in each coding circuit may be the same as ISO/IEC JTC1/SC29/WG11 N1825 described in the above-described conventional technique, and a decoding time stamp and a composite time stamp are used.

[0047] The decoding time stamp is used for only an interpolative predicted picture, and only the composite time stamp is used for video, audio and scene data of the other prediction modes. That is, the decoding timing and the timing at which the decoding data is allowed to be used by the composition circuit 5 are assumed to be equal to each other. However, it is important that a fixed delay is set between the coding apparatus or a storage medium and the decoding apparatus, and the decoding of the decoding apparatus may be terminated after a fixed time elapses from the time represented by the time stamp.

BNSDOCID: <EP_____0924934A1_I_>

[0048] The composition circuit 5 receives the audio signal output from the audio coding circuit 1, the video signal output from the video coding circuit 2 and the scene data output from the scene coding circuit 4 to compose a scene according to a scene description described in the scene data, and outputs a composite picture, the audio signal and the time stamp representing the composition timing. This time stamp is not shown in ISO/IEC JTC1/SC29/WG11 N1825, and in this specification, it is called as "display time stamp". That is, the composition timing and the display timing are assumed to be equal to each other. However, it is important that a fixed delay is set between the coding apparatus or the storage medium and the decoding apparatus, and the composition of the decoding apparatus may be terminated after a fixed time elapses from the time represented by the time stamp.

[0049] The multiplexing circuit 6 receives both of the compressed data and the time stamp representing the decoding timing which are output from the audio coding circuit 1, both of the compressed data and the time stamp representing the decoding timing which are output from the video coding circuit 2, both of the compressed data and the time stamp representing the decoding timing which are output from the scene coding circuit 4, the time stamp representing the composition timing which is output from the composition circuit 5, and clocks supplied from a clock generating circuit 8 described later, and generates and outputs a bit stream.

[0050] The display circuit 7 receives the composite picture signal and the audio signal which are output from the composition circuit 5, and display/reproduces the data through a display for video data and through a speaker or the like for audio data. The clock generating circuit 8 generates clocks as clock inputs (CLK) to the audio coding circuit 1, the video coding circuit 2, the scene coding circuit 4, the composition circuit 5, and the multiplexing circuit 6.

[0051] Fig. 2 shows the construction of the audio coding circuit 1, the video coding circuit 2 and the scene coding circuit 4. The input signals to the respective coding circuits are different from one another, however, the respective coding circuits have the functionally common structure which comprises encoder 11, decoder 12, memory 13, buffer 14 and buffer 15. The encoder 11 receives the input signal and locally decoded data supplied from the memory 13 (described later) and outputs the compressed data. Further, it outputs the time stamp representing the decoding timing. For example, it outputs the time at which the coding is finished. The decoder 12 receives the compressed data output from the encoder 11 and the locally decoded data supplied from the memory 13 and outputs new locally decoded data. The memory 13 stores the locally decoded data supplied from the decoder 12, and outputs the data to the encoder 11 and the composition circuit 5. The buffer 14 buffers the time stamp representing the decoding timing supplied from the encoder 11, and outputs it to the multiplexing circuit 6. The buffer 15 buffers the compressed data output from the encoder 11, and outputs the data to the multiplexing circuit 6. Further, clocks are supplied from the clock generating circuit 8, and these clocks are set as clock inputs (CLK) to the encoder 11 and the decoder 12.

[0052] In Fig. 2, the locally decoded data stored in the memory 13 are used as an input to the encoder 11 and the decoder 12 for a subsequent coding process. However, these data may not be used for the subsequent coding process in such a case as coding of a still picture.

[0053] Fig. 3 shows the construction of the composition circuit 5 of Fig. 1. The composition circuit 5 comprises scene generating circuit 201, buffer 202, conversion processing circuit 203, texture generating circuit 204, raster circuit 205, delay circuit 206 and frame buffer 207.

[0054] The scene generating circuit 201 receives the scene data from the scene coding circuit 4 to generate a scene graph, and outputs a scene drawing command and intermediate data together with a time stamp representing the composition timing. In the case of a two-dimensional scene, coordinate data, graphics data, text data are generated at every object in a scene. Further, the fore-and-aft relationship of respective objects is added. In the case of a three-dimensional scene, setting of a camera, setting of the angle of field of view, setting of alight source, deletion of objects out of the visual field are further performed. The buffer 202 buffers the time stamp representing the composition timing which is supplied from the scene generating circuit 201.

[0055] The conversion processing circuit 203 receives a scene drawing command and intermediate data supplied from the scene generating circuit 201 to execute conversion processing such as coordinate transformation, light-source calculation, clipping and outputs new intermediate data. Further, it receives a texture from a texture generating circuit 204 described later, and maps it into an object in a scene. In the case of the two-dimensional scene, movement, rotation, enlargement, reduction of object, and other processing are carried out. In the case of the three-dimensional scene, the effect of the light source, and hidden surface algorithm in depth direction are further carried out. Through the above processing, the position information and the color information of each object in a scene that is viewed from a current visual point are determined and output.

[0056] The texture generating circuit 204 receives the video data supplied from the video coding circuit 2, the drawing command supplied from the scene generating circuit 201 and the coordinate information supplied from the conversion processing circuit 203, deforms into a texture the video data which are mapped into an object in a scene, and then outputs the texture thus obtained. The present invention is based on the assumption that the scene composition is repeated every frame, and thus it is general that the video data corresponds to one

BNSDOCID: <EP_____0924934A1_I_>

picture.

[0057] The raster circuit 205 receives the intermediate data from the conversion processing circuit 203 to convert the intermediate data to raster data on a pixel basis. The delay circuit 206 receives the audio data from the audio coding circuit 1 to delay the audio data in consideration of the time lapse of the processing executed from the scene generating circuit 201 to the raster circuit 205; and outputs the audio data thus delayed to the display circuit 7. The frame buffer 207 stocks the raster data supplied from the raster circuit 205, and outputs the raster data thus stored to the display circuit 7. The scene generating circuit 201, the conversion processing circuit 203, the texture generating circuit 204 and the raster circuit 205 are supplied with the clocks (CLK) from the clock generating circuit 8.

[0058] Fig. 45 shows another embodiment of the composition circuit 5 of Fig. 1, and the composition circuit 5 comprises interface circuit 21, central processing unit (CPU) 22, conversion processing circuit 23, raster circuit 24, texture generating circuit 25, frame buffer 26, delay circuit 27, counter 28 and memory 29. The respective circuits are connected to one another through a bus.

[0059] The interface circuit 21 receives the audio data supplied from the audio coding circuit 1, the video data supplied from the video coding circuit 2 and the scene data supplied from the scene coding circuit 4 and outputs the time stamp representing the composition timing described later to the multiplexing circuit 6. That is, it serves as an interface between each circuit connected to the bus and the external.

[0060] CPU 22 performs various software processing such as initial-stage processing needed for scene composition, generation of a scene graph on the basis of the scene data supplied from the scene coding circuit 4, allocation of an operation to each circuit on the basis of analysis of the scene graph, a schedule management of each more general circuit resource. Further, it outputs the time stamp representing the composition timing to the interface circuit 21, and performs an emulation of operation frequency control by using a clock calculation value given from the counter 28 described later.

[0061] The conversion processing circuit 23 performs the same processing as the conversion processing circuit 203 shown in Fig. 3 in response to the drawing command from the CPU 22. The raster circuit 24 performs the same processing as the raster circuit 205 of Fig. 3 in response to the drawing command from the CPU 22. The raster data thus finally obtained are written into the frame buffer 26 described later. The texture generating circuit 25 performs the same processing as the texture generating circuit 204 of Fig. 3 in response to the drawing command from the CPU 22. The frame buffer 26 stores the raster data obtained from the raster circuit 24 and outputs the data thus stored to the display circuit 7. The delay circuit 27 delays the audio signal from the audio coding circuit 1 in consideration of the calculation time for a series of composition processing, and outputs the audio signal thus delayed to the display circuit 7. The counter 28 counts the number of clocks supplied from the clock generating circuit 8, and outputs the count number to the CPU 22 as occasion demands.

[0062] In this case, the operation frequency of the CPU 22, the conversion processing circuit 23, the raster circuit 24 and the texture circuit 25 is given from another clock generating circuit. However, the clocks supplied from the clock generating circuit 8 may be used. The memory 29 is used to store control data and intermediate data needed for the calculation in each of the CPU 22, the conversion processing circuit 23, the raster circuit 24 and the texture generating circuit 25.

[0063] Fig. 4 is a diagram showing the construction of the multiplexing circuit 6 of Fig. 1, and the multiplexing circuit 6 comprises multiplexer 31, counter 32, additive information holding circuit 33, and buffer 34. The multiplexer 31 multiplexes the compressed data of the audio signal and the time stamp representing the decoding timing which are supplied from the audio coding circuit 1, the compressed data of the video signal and the time stamp representing the decoding timing which are supplied from the video coding circuit 2, the compressed data of the scene data and the time stamp representing the decoding timing which are supplied from the scene coding circuit 4, the time stamp representing the composition timing supplied from the composition circuit 5, a clock count value supplied from the counter 32 described later, and additive information supplied from the additive information holding circuit 33 described later, and generates and outputs a bit stream.

[0064] The counter 32 counts the clocks supplied from the clock generating circuit 8, and outputs the count number. The additive information holding circuit 33 holds overhead information that is preset to be added for generation of a bit stream, and outputs the overhead information. The buffer 34 buffers the bit stream output from the multiplexer 31 and outputs the bit stream. The buffer 34 is needed when the present invention is applied to a transmission system, however, it is not necessarily required when the present invention is applied to a storage system.

[0065] Next, the operation of the coding apparatus according to the present invention will be described with reference to Figs. 1 to 4 and Fig. 45.

[0066] Each of the audio coding circuit 1, the video coding circuit 2 and the scene coding circuit 4 performs compression coding on the input signal thereto, and also outputs the time stamp representing the decoding timing. As shown in Fig. 2, the encoder 11 first performs compression processing by using the input signal and the locally-decoded data output from the memory 13, and writes the compressed data into the buffer 15. At the same time, the encoder 11 outputs the time stamp representing the decoding timing, and writes the time stamp into the buffer 14. Subsequently, the decoder 12 decodes the compressed data supplied from the

BNSDOCID: <EP_____0924934A1_I_>

encoder 11, and adds the compressed data thus decoded to the locally-decoded data supplied from the memory 13 to create new locally-decoded data. This locally-decoded data is newly written into the memory 13.

[0067] The interface circuit 3 to the scene data supports various input modes for scene design and scene update such as a keyboard input, a mouse input, and it converts input data to coherent scene data and outputs the data thus obtained to the scene coding circuit 4. With respect to specific scene data, use of data replacement and data differential may be considered as in the case of the concepts of the intra-frame coding, inter-frame coding of video signals. The switching between the data replacement and the data differential is managed by the scene coding circuit 4 in response to an instruction from the interface circuit 3. Since VRML is originally text data, there may be considered a mode in which compression isn't performed and scene data are directly transmitted.

[0068] The composition circuit 5 performs the scene composition by using the audio data obtained from the audio coding circuit 1, the video data obtained from the video coding circuit 2 and the scene data obtained from the scene coding circuit 4. At the same time, it outputs the time stamp representing the composition timing. In this case, each of the data is directly used the locally-decoded data stored in the memory of the coding circuit. More specifically, as shown in Fig. 3, the scene generating circuit 201 creates a scene graph on the basis of the scene data supplied from the scene coding circuit 4, and outputs the scene drawing command and the intermediate data. At this time, it outputs the time stamp representing the composition timing at the same time, and writes it into the buffer 202. Subsequently, the conversion processing circuit 203 executes the above conversion processing on the basis of the drawing command from the scene generating circuit 201, and outputs the coordinate information and the color information of an object.

[0069] Further, the texture data supplied from the texture generating circuit 204 are mapped into an object in a scene. In parallel to the processing, the texture generating circuit 204 deforms the video data obtained from the video coding circuit 2 on the basis of the drawing command supplied from the scene generating circuit 201 and the coordinate information supplied from the conversion processing circuit 203. The conversion processing circuit 203 and the texture generating circuit 204 execute the respective processing while communicating data therebetween.

[0070] Subsequently, the raster circuit 205 converts the data from the conversion processing circuit 203 to raster data on a pixel basis on the basis of the coordinate information and the color information of the object which are supplied from the conversion processing circuit 203, and writes the conversion result into the frame buffer 207. The audio signal supplied from the audio coding circuit 1 is delayed and output by the delay circuit 206. The same operation is also carried out in the construction of Fig. 45. In this case, not only the audio signal is delayed, but also a special effect and other effects can be easily implemented by CPU 22.

[0071] There is a case where an event dependent on time is described in the scene data. This event is classified into a continuous event which varies on time axis, and a discrete event which is one-shot event on time axis. With respect to processing of these events, the continuous event is processed as an event occurring at the time stamp representing the composition timing, and the discrete event is processed as an event occurring at the time when the time stamp representing the composition timing passes the generation time of the discrete even. Accordingly, when the same event processing is carried out according to the time stamp representing the composition timing at the reception side, it is ensured that the same composition result can be implemented at both the transmission side and the reception side.

[0072] The specific processing is carried out by the scene generating circuit 201 of Fig. 3 or the CPU 22 of Fig. 45. Therefore, the scene generating circuit 202 or the CPU 22 has a counter or the like at the inside or the outside thereof for time management. The counter is set to zero at the time when a session is started, and it is driven with clocks supplied by the clock generating circuit 8 in the case of the scene generating circuit 202, while it is driven with clocks which exist independently of the clock generating circuit 8 in the case of the CPU 22.

[0073] The multiplexing circuit 6 multiplexes the compressed data, the time stamp and the reference clock value to generate a bit stream. More specifically, as shown in Fig. 4, in accordance with a predetermined timing, the multiplexer 31 multiplexes the compressed data and the time stamp supplied from the audio coding circuit 1, the compressed data and the time stamp supplied from the video coding circuit 2, the compressed data and the time stamp supplied from the scene coding circuit 4, the time stamp supplied from the composition circuit 5, the count value of the clocks supplied from the counter 32 and an overhead representing system information supplied from the additive information holding circuit 33.

[0074] The counter 32 counts the clocks supplied from the clock generating circuit 8, and outputs the count value thereof. The additive information holding circuit 33 holds not only the overhead representing the system information, but also multiplexing management information such as the bit length of each data to be multiplexed, the time stamp and supplies the information as control information to the multiplexer 31. As a specific mode of the additive information holding circuit may be used ROM containing predetermined fixed data, a ROM card or RAM into which data are loaded at an initialization time through a keyboard or the like.

[0075] Fig. 26 shows a finally-obtained bit stream.

BNSDOCID: <EP_____0924934A1_I_>

That is, the bit stream comprises the reference clock value, and the time stamps and compressed data for audio, video, scene data respectively. Each time stamp representing the decoding timing is appended to the corresponding compressed data, and the time stamp representing the composition timing is selectively appended to the compressed video data, to the compressed scene data, or out of the compressed data as in the case of the reference clock.

[0076] The display circuit 7 performs display and reproduction of the composite picture signal and the audio signal supplied from the composition circuit 5, whereby a transmitter can observe, on the spot, a picture desired to be composed by itself and the audio signal thereof. Further, the scene can be suitably updated through the interface circuit 3. The clock generating circuit 8 continues to generate clocks (CLK) in a coherent way, and supplies the clocks thus generated to the audio coding circuit 1, the video coding circuit 2, the scene coding circuit 4, the composition circuit 5 and the multiplexing circuit 6.

[0077] In the coding apparatus of the first embodiment according to the present invention, no consideration is given to the delay needed to the composition processing. That is, when all the processing is carried out while the frame rates thereof are kept within given limits, the time chart representing the processing flow for coding, decoding and composition at the coding apparatus side is shown in Fig. 28. Here, the coding corresponds to the processing of the encoder in the coding circuit, and the decoding corresponds to the processing of the decoder in the coding circuit, that is, the creation of the locally-decoded data. The composition corresponds to the processing of the composition circuit. The time period from the start time of a coding operation to the start time of the next coding operation corresponds to the frame rate of the input video signal. Further, the time period from the start time of a composition operation to the start time of the next composition operation corresponds to the frame rate of the composite picture. In Fig. 28, the coding, the decoding and the composition are expressed as sequential processing. However, by dividing each of the coding and decoding operations into plural threads, the parallel processing on plural signals may be supported. An example of the occurrence timing of the decoding time stamp and the composition time stamp is shown in Fig. 28. However, for the purpose of keeping a fixed delay between the coding apparatus side and the decoding apparatus side, the occurrence timing may be set to the time when the decoding, composition are terminated, or to any time. In this case, the coding and the decoding are assumed to be absolutely finished within one frame period.

[0078] On the other hand, when the composition time is needed to be long, there is a case where it is required to continue the composition operation until the time of the next composition frame as shown in Fig. 29. When the parallel processing of the coding/decoding and the composition is not supported, or when the coding/decoding and the composition cannot be executed in parallel due to a competition problem of an access to the memory for storing the locally-decoded data, it is difficult to continue the composition or the coding/decoding no longer.

[0079] As a countermeasure to the above case, by performing the coding, decoding and composing processing through the time chart of Fig. 30, the coding/decoding can be continued. That is, when the composition processing is not terminated until the time set at the coding apparatus side, the coding/decoding processing of the video frame at that time is paused, and the extra time corresponding to the pause time is allocated to the composition processing. For the video data of the paused frame, nothing (containing the time stamp) is transmitted, or the coding is performed on the assumption that there is no variation between the frame concerned and the preceding frame. After the composition of the frame concerned is terminated, a next composition operation is started in accordance with the frame rate of the composite picture. When the composition concerned is continued until this time point, the composition circuit itself pauses for the next composition. However, the coding operation is not paused because if the coding of the audio signal is paused, sound quality would be remarkably reduced due to occurrence of missed sections.

[0080] Fig. 31 is a time chart for the coding, the decoding and the composition when the coding/decoding for plural input signals is carried out. The coding/decoding operation is sequentially carried out on two input signals, and then the composition processing is carried out. The decoding time stamp and the composition time stamp are generated as shown in Fig. 31, respectively.

[0081] Fig. 32 is a time chart when the composition processing is continued until a firs input signal of a next frame. In this case, as shown in Fig. 33, the coding/decoding processing of the first input signal is paused, and for the video data of the paused frame, nothing (containing the time stamp) is transmitted, or the coding is performed on the assumption that there is no variation between the frame concerned and the preceding frame. For a second input signal, the coding/decoding is carried out, and the composition is carried out.

[0082] Likewise, Fig. 34 is a time chart when the composition processing is continued until the second input signal of the next frame. In this case, as shown in Fig. 35, the coding/decoding of the first input signal and the coding/decoding of the second input signal are paused. For the video data of the paused frame, nothing (containing the time stamp) is transmitted, or the coding is carried out on the assumption that there is no variation between the frame concerned and the preceding frame.

[0083] When the composition processing concerned is not finished until the time when the next composition

BNSDOCID: <EP_____0924934A1_I_>

processing is carried out, the composition circuit itself pauses for the next composition processing. In the decoding apparatus, the decoding and composition operations are carried out in accordance with the time stamp in the bit stream, and thus when no decoding time stamp exists, the decoding processing is automatically skipped. Therefore, the frame rate of the video signal is temporarily reduced, however, the composition processing is stably performed.

[0084] Fig. 5 is a block diagram showing a second embodiment of the coding apparatus which is designed so that the coding/decoding can be continued even in the case where the continuity of the composition is requested until the time of the next composite frame.

[0085] In this embodiment, the coding apparatus of the first embodiment is newly added to scheduling circuit 153. That is, the video coding circuit 151 is added to a control line extending from the scheduling circuit 153 in addition to the construction of the video coding circuit 2 of Fig. 1. In addition to the construction of the composition circuit 5 of Fig. 1, a composition circuit 152 is designed so as to output a signal representing the composition status, that is, whether the composition is terminated or not, to the scheduling circuit 153. Upon receiving the composition status signal from the composition circuit 152, the scheduling circuit 153 controls the operation of the coding circuit 151.

[0086] Fig. 6 shows the construction of the coding circuit 151, and the encoder 11 and the decoder 12 of Fig. 2 are replaced by an encoder 154 and a decoder 155, respectively. The coding operation of the encoder 154 and the decoding operation of the decoder 155 are together controlled on the basis of the input from the scheduling circuit 153.

[0087] Fig. 7 shows a first embodiment of the composition circuit 152 of Fig. 5, and it is designed in such a way that the scene generating circuit 201, the conversion processing circuit 203, the texture generating circuit 204 and the raster circuit 205 of Fig.3 are replaced by a scene generating circuit 211, a conversion processing circuit 212, a texture generating circuit 213 and a raster circuit 214, and an OR circuit 215 is newly added. Each of the scene generating circuit 211, the conversion processing circuit 212, the texture generating circuit 213 and the raster circuit 214 has an output representing whether the processing thereof is terminated or not, in addition to the construction of each of the scene generating circuit 201, the conversion processing circuit 203, the texture generating circuit 204 and the raster circuit 205 of Fig.3.

[0088] The OR circuit 215 receives the status inputs from the scene generating circuit 211, the conversion processing circuit 212, the texture generating circuit 213 and the raster circuit 214 to perform OR operation between the status inputs thus received, and outputs the OR-operation result. In this case, it is assumed that "1" is set under processing and "0" is set at the termination of the processing.

[0089] Fig. 46 shows a second embodiment of the composition circuit 152, and it is constructed so that the interface circuit 21 of Fig. 45 is replace by an interface circuit 156. In addition to the construction of the interface circuit 21, the interface circuit 156 has an output representing the composition status of the composition circuit 152 to the scheduling circuit 153.

[0090] Next, the operation of the second embodiment of the coding apparatus according to the present invention will be described with reference to Figs. 5 to 7 and Fig. 46. The basic operation of the coding operation is the same as that of the circuit of Fig. 1. However, a signal representing the composition status is transmitted from the composition circuit 152 to the scheduling circuit 153. As the signal representing the composition status, "1" is output when any one or more of the scene generating circuit 211, the conversion processing circuit 212, the texture generating circuit 213 and the raster circuit 214 are under operation, and "0" is output when all of the circuits are at rest as shown in Fig. 7.

[0091] In the construction of Fig. 46, the CPU 22 transmits the same signal to the scheduling circuit 153 through the interface circuit 156. Upon receiving the signal, the scheduling circuit 153 outputs "1" when the input signal is "1", and outputs "0" when the input signal is "0". As shown in Fig. 6, the encoder 154/decoder 155 receives this signal, and the coding circuit 151 does not start the coding/decoding even at a predetermined timing when the input signal is "1" while the coding circuit 151 starts the coding/decoding when the input signal is "0".

[0092] In Fig. 28, the coding/decoding is illustrated as being sequentially carried out, and there occurs a problem in existence of decoding data when the input signal is set to "1" at the coding start time and to "0" at the decoding start time. However, this problem could be avoided by presetting the decoding operation so that the decoding operation is not carried out when the input signal is "1" at the coding start time.

[0093] The problem of the composition processing time shown in Fig. 29 can be also avoided by scheduling the coding operation, the decoding operation and the composition operation as shown in Fig. 36. In this case, when the composition has not been terminated until the coding start timing of the next frame which is set by the coding apparatus, the coding/decoding is not paused, but the composition is paused, and then the composition is resumed at the time when the coding/decoding is finished. When the composition concerned has not been terminated until the next coding start timing, the composition is paused again, and the composition processing is on standby until the coding/decoding is finished.

[0094] In the decoding apparatus, the decoding and the composition are carried out in response to the time stamp in the bit stream, and thus it is settled that in response to the decoding time stamp, the decoding is started while the composition is paused, and the com-

BNSDOCID: <EP_____0924934A1_I_>

position is resumed at the time when the decoding is finished. Accordingly, the frame rate of the composite picture is temporarily reduced, however, the coding of the video signal based on a fixed frame rate is expected. This is effective when only the compressed data of the video signal is afterwards reused for edition or the like.

[0095] Fig. 37 is a diagram showing a countermeasure based on the scheduling of the coding, the decoding and the composition for plural input signals of Figs. 32 and 34. Basically, the same countermeasure as shown in Fig. 36 is taken.

[0096] Fig. 8 is a block diagram showing a third embodiment of the coding apparatus according to the present invention in which when the coding/decoding operation is enabled to continue by pausing the composition operation in the case where the continuity of the composition until the time of a next composite frame is requested.

[0097] In this embodiment; a scheduling circuit 165 is newly added to the coding apparatus of the first embodiment. An audio coding circuit 161, a video coding circuit 162 and a scene coding circuit 163 has the same construction as the audio coding circuit 1, the video coding circuit 2 and the scene coding circuit 4 of Fig. 1 respectively, and also each of the circuit is further designed to output to the scheduling circuit 165 a signal representing a coding status, that is, whether the coding is carried out or not.

[0098] In addition to the construction of the composition circuit 5 of Fig. 1, the composition circuit 164 is added with a control line extending from the scheduling circuit 165. The scheduling circuit 165 receives the status inputs from the coding circuit 161, the coding circuit 162 and the coding circuit 163 to control the operation of the composition circuit 164.

[0099] Fig. 9 shows the construction of the coding circuits 161, 162 and 163, and the encoder 11 and the decoder 12 of Fig. 2 are replaced by encoder 166 and decoder 167. Further, OR circuit 168 is newly provided. In addition to the construction of the encoder 11, the decoder 12, each of the encoder 166 and the decoder 167 is further designed so as to output to the OR circuit 168 a signal representing whether the processing thereof is finished or not. The OR circuit 168 receives the status inputs from the encoder 166 and the decoder 167, and outputs the OR output to the scheduling circuit 165. In this case, it is assumed that "1" is set under processing, and "0" is set at the time when the processing is finished.

[0100] Fig. 10 shows a first embodiment of the composition circuit 164 of Fig. 8. The scene generating circuit 201, the conversion processing circuit 203, the texture generating circuit 204 and the raster circuit 205 of Fig. 3 are replaced by scene generating circuit 221, conversion processing circuit 222, texture generating circuit 223 and raster circuit 224, and further control circuit 225 is newly added. In addition to the construction of each of the scene generating circuit 201, the conversion processing circuit 203, the texture generating circuit 204 and the raster circuit 205 of Fig.3, each of the scene generating circuit 221, the conversion processing circuit 222, the texture generating circuit 223 and the raster circuit 224 is further provided with an input line from the control circuit 225. The control circuit 225 receives an input from the scheduling circuit 165 and outputs it to each of the scene generating circuit 221, the conversion processing circuit 222, the texture generating circuit 223 and the raster circuit 224 to control the operation of each circuit.

[0101] Fig. 47 shows a second embodiment of the composition circuit 164, and in this embodiment the interface circuit 21 of Fig. 45 is replaced by an interface circuit 169. In addition to the construction of the interface circuit 21, the interface circuit 169 is designed so as to receive an input from the scheduling circuit 165.

[0102] The operation of the third embodiment of the coding apparatus of the present invention will be described with reference to Figs. 8 to 10 and Fig. 47. The basic operation of the coding operation is the same as the circuit of Fig. 1. However, each of the audio coding circuit 161, the video coding circuit 162 and the scene coding circuit 163 transmits the coding status to the scheduling circuit 165. In the coding circuit 161, the coding circuit 162 and the coding circuit 163, an encoder 166 and a decoder 167 output a coding state and a decoding state to the OR circuit 168 respectively as shown in Fig. 9. The output signal is set to "1" when the encoder (decoder) is under operation, and "0" when it is at a rest. Therefore, the output of the OR circuit 168 is set to "1" when either of the encoder and the decoder is under operation, and "0" when both the encoder and the decoder are at a rest.

[0103] The scheduling circuit 165 receives inputs from the coding circuits 161 to 163 to perform OR operation therebetween, and outputs the OR result. In the composition circuit 164, the control circuit 225 receives an input from the scheduling circuit 165 and outputs it to the scene generating circuit 221, the conversion processing circuit 222, the texture generating circuit 223 and the raster circuit 224 as shown in Fig. 10. At the time when the input value from the control circuit 225 varies from "0" to "1", each of the scene generating circuit 221, the conversion processing circuit 222, the texture generating circuit 223 and the raster circuit 224 stores intermediate data and pauses the processing thereof. At the time when the input value varies from "1" to "0", each circuit recovers the intermediate data and resumes the processing. When the input value is equal to "1" at all times, each circuit is at a rest. When the input value is equal to "0" at all times, the processing is started in synchronism with the composition timing.

[0104] In the first to third embodiments of the coding apparatus according to the present invention, the same clocks are supplied from the same clock generating circuit for the audio signal, the video signal and the scene data. However, according to the system shown in

BNSDOCID: <EP_____0924934A1_I_>

ISO/IEC JTC1/SC29/WG11 N1825 described in the conventional technique, it is allowed that different clocks may be provided for each of the audio signal, the video signal and the scene data. Accordingly, in the coding apparatus of the present invention, there may be provided different clocks between the audio signal, the video signal and the scene data.

[0105] Fig. 11 shows a fourth embodiment of the coding apparatus according to the present invention. In the fourth embodiment, a clock generating circuit is individually provided to each of the audio coding circuit 1, the video coding circuit 2, the scene coding circuit 4 and the composition circuit 5 in the first embodiment. That is, in place of the clock generating circuit 8 of Fig. 1, three clock generating circuits 171, 172 and 173 are provided. The audio coding circuit 1 is supplied with clocks (CLK1) from the clock generating circuit 171, the video coding circuit 2 is supplied with clocks (CLK2) from the clock generating circuit 172 and the scene coding circuit 4 and the composition circuit 5 are supplied with clocks (CLK3) from the clock generating circuit 173.

[0106] In addition to the construction of the multiplexing circuit 6 of Fig. 1, the multiplexing circuit 174 is designed to receive clock inputs from three clock generating circuits 171, 172, 173.

[0107] Fig. 12 shows the construction of the multiplexing circuit 174 of Fig. 11. The multiplexing circuit 174 has three counters 32 in association with the three clock generating circuits 171, 172 and 173 in addition to the construction of the multiplexing circuit 6 of Fig. 4. A multiplexer 175 is designed so as to receive and multiplex inputs from the three counters 32 in addition to the construction of the multiplexer 31 of Fig. 4.

[0108] Next, the operation of the fourth embodiment of the coding apparatus according to the present invention will be described with reference to Figs. 11 to 13. The basic operation of the coding is the same as the circuit of Fig. 1. The difference from the circuit of Fig. 1 resides in that the audio coding circuit 1, the video coding circuit 2 and both the scene coding circuit 4 and the composition circuit 5 are respectively operated with the respective clocks supplied from the three different clock generating circuits 171, 172 and 173, and that the multiplexing circuit 174 multiplexes the clocks supplied from the three different clock generating circuits 171, 172 and 173.

[0109] The final bit stream is shown in (1) of Fig. 27. That is, the bit stream comprises a reference clock value, a time stamp and compressed data for each of audio, video and scene data. Each time stamp representing the decoding timing is appended to the corresponding compressed data, and the time stamp representing the composition timing is appended to the compressed scene data which is an output of the scene coding circuit 4 operating with the same clock as the composition circuit 5.

[0110] Fig. 13 shows a fifth embodiment of the coding apparatus of the present invention. According to the coding apparatus of this embodiment, three different clock generating circuits 171, 172 and 173 are respectively allocated to the audio coding circuit 1, the video coding circuit 151, and both the scene coding circuit 4 and the composition circuit 152 in the coding apparatus of the second embodiment. The multiplexing circuit 174 has the same construction as the fourth embodiment.

[0111] Fig. 14 shows a sixth embodiment of the coding apparatus according to the present invention. According to the coding apparatus of this embodiment, three different clock generating circuits 171, 172 and 173 are respectively allocated to the audio coding circuit 161, the video coding circuit 162, and both the scene coding circuit 163 and the composition circuit 164 in the coding apparatus of the third embodiment. The multiplexing circuit 174 has the same construction as the fourth embodiment.

[0112] Fig. 15 shows a seventh embodiment of the coding apparatus of the present invention. According to the seventh embodiment, three different clock generating circuits 171, 172 and 173 are respectively allocated to the audio coding circuit 1, both the video coding circuit 2 and the composition circuit 5, and the scene coding circuit 4 in the coding apparatus of the first embodiment. The multiplexing circuit 174 has the same construction as the fourth embodiment. The basic operation of the coding is as the same as the circuit of Fig. 1. The difference from the circuit of Fig. 1 resides in that the audio coding circuit, both the video coding circuit 2 and the composition circuit 5, and the scene coding circuit 4 are operated with the respective clocks supplied from the different three clock generating circuits 171, 172 and 173, and that the multiplexing circuit multiplexes the clocks supplied from the three different clock generating circuits 171, 172 and 173.

[0113] The final bit stream is shown in (2) of Fig. 27. That is, the bit stream comprises a reference clock value, a time stamp and compressed data for each of audio, video and scene data. Each time stamp representing the decoding timing is appended to the corresponding compressed data, and the time stamp representing the composition timing is appended to the compressed video data which is an output of the video coding circuit 2 operating with the same clocks as the composition circuit 5.

[0114] Fig. 16 shows an eighth embodiment of the coding apparatus according to the present invention. According to the eight embodiment, three different clock generating circuits 171,172, and 173 are respectively allocated to the audio signal circuit 1, both the video signal circuit 15 and the composition circuit 5, and the scene coding circuit 4 in the coding apparatus of the second embodiment. The multiplexing circuit 174 has the same construction as the fourth embodiment.

[0115] Fig. 17 shows a ninth embodiment of the coding apparatus according to the present invention. According to the ninth embodiment, three different clock generating circuits 171, 172 and 173 are respectively

BNSDOCID: <EP_____0924934A1_I_>

allocated to the audio coding circuit 161, and both the video coding circuit 162 and the composition circuit 164, and the scene coding circuit 163 in the coding apparatus of the third embodiment of the present invention. The multiplexing circuit 174 has the same construction as the fourth embodiment.

[0116] Fig. 18 is a block diagram showing a first embodiment to the decoding apparatus of the present invention. The decoding apparatus of the present invention comprises a separation circuit (demultiplexing circuit) 41, a decoding circuit 42 for audio signals (hereinafter referred to as "audio decoding circuit"), a decoding circuit 43 for video signals (hereinafter referred to as "video decoding circuit"), a decoding circuit 44 for scene data (hereinafter referred to as "scene decoding circuit"), a composition circuit 45, a display circuit 46, a clock generating circuit 47 and an interaction circuit 48.

[0117] The separation circuit 41 outputs from an input bit stream the compressed data and the time stamp representing the decoding timing for the audio signal, the compressed data and the time stamp representing the decoding timing for the video signal, the compressed data and the time stamp for the scene data, the time stamp representing the composition timing and a reference clock value supplied to the clock generating circuit 47 (described later).

[0118] The audio decoding circuit 42 decodes the compressed data input from the separation circuit 41 at the time represented by the time stamp representing the decoding timing which is input from the separation circuit 41. The video decoding circuit 43 decodes the compressed data input from the separation circuit 41 at the time represented by the time stamp representing the decoding timing which is input from the separation circuit 41. The scene decoding circuit 44 decodes the compressed data input from the separation circuit 41 at the time represented by the time stamp representing the decoding timing which is input from the separation circuit 41.

[0119] The composition circuit 45 performs the composition processing on the audio signal from the audio decoding circuit 42, the video signal from the video decoding circuit 43 and the scene data from the scene decoding circuit 44 input thereto in accordance with a scene description described in the scene data at the time represented by the time stamp representing the composition timing input from the separation circuit 41, and outputs a composite picture and the audio signal. Further, it accepts input data from the interaction circuit 48 described later to implement user interaction such as movement of a viewing point.

[0120] The display circuit 46 receives the composite picture signal and the audio signal from the composition circuit 45, and displays/reproduces these signals through a display or the like for pictures and through a speaker or the like for sounds. The clock generating circuit 47 generates clocks (CLK10) in accordance with the reference clock value supplied from the separation circuit 41, and supplies the clocks to the audio decoding circuit 42, the video decoding circuit 43, the scene decoding circuit 44 and the composition circuit 45. The clock generating circuit 47 is generally constructed as PLL (Phased Locked Loop), and the reference clock value is used to control the oscillation frequency of the clocks.

[0121] The interaction circuit 48 accepts an interaction such as a keyboard input, a mouse input or the like from a viewer to convert it to data representing movement of a viewing point or the like, and outputs the conversion result to the composition circuit 45.

[0122] Fig. 19 shows the construction of the separation circuit 41 of Fig. 18, and it comprises buffer 51, demultiplexer 52 and additive information holding circuit 53. The buffer 51 buffers a bit stream which is transmitted through a network or read out from a storage medium such as a disk or the like. The demultiplexer 52 separates the bit stream input from the buffer 51 into the compressed data and the time stamp representing the decoding timing for the audio information, the compressed data and the time stamp representing the decoding timing for the video information, the compressed data and the time stamp representing the decoding timing for the scene data, the time stamp representing the composition timing, the reference clock value and overhead serving as system information on the basis of the management information such as bit length which are hold in the additive information holding circuit 53.

[0123] The additive information holding circuit 53 holds not only the overhead representing the system information, but also the multiplexing management information such as the bit length of each data to be multiplexed, the time stamps and supplies these data as control information to the demultiplexer 52. As specific modes of the additive information holding circuit 53 may be considered a ROM containing predetermined fixed data, a ROM card, a RAM into which data are loaded through a keyboard or the like at an initialization time, a RAM for storing bit stream information contained in the overhead serving as the system information in the bit stream or the like.

[0124] Fig. 20 shows the construction of the decoding circuits 42, 43 and 44 of Fig. 18, and it comprises a buffer 61, a buffer 62, a decoder 63 and a memory 64. The buffer 61 buffers a time stamp representing a decoding timing which is supplied from the separation circuit 41. The buffer 62 buffers a compressed data which is supplied from the separation circuit 41. The decoder 63 receives the compressed data supplied from the buffer 62 and the decoding data supplied from a memory 64 described later at the time of the time stamp representing the decoding timing supplied from the buffer 61 to perform the decoding operation. The decoder 63 is supplied with clocks from the clock generating circuit 47.

BNSDOCID: <EP_____0924934A1_I_>

[0125] The memory 64 stores the decoding data supplied from the decoder 63. In this construction, the decoding operation of the decoder 63 is carried out on the assumption that the decoding data stored in the memory 64 are used. However, there is a case where the decoding data are not used as in the case of an intra-frame coding of video. In the case of scene data, text data that are not compressed may be considered. In this case, the data are merely written into the memory modification.

[0126] Fig. 21 shows 64 with no a first embodiment of the composition circuit 45 of Fig. 18. According to this embodiment, in the construction of Fig. 3, the scene generating circuit 201 is replaced by a scene generating circuit 231 and the buffer 202 is replaced by a buffer 232, and a buffer 233 is further added. The scene generating circuit 231 is designed so that the output line of the time stamp representing the composition timing is removed from the scene generating circuit 201 and in place of the output line thus removed, input lines from the buffer 232 and the buffer 233 are added. The buffer 232 buffers the time stamp representing the composition timing from the separation circuit 41. The buffer 233 buffers interaction data from the interaction circuit 48. The clocks from the clock generating circuit 47 are supplied to the scene generating circuit 231, the conversion processing circuit 203, the texture generating circuit 204 and the raster circuit 205.

[0127] Fig. 48 shows a second embodiment of the composition circuit 45. In the construction of Fig. 48, the interface circuit 21 of Fig.45 is replaced by an interface circuit 49. The interface circuit 49 is designed so that the output line to the multiplexing circuit 6 is removed from the interface circuit 21 of Fig. 45, and in place of the output line thus removed an input line for the time stamp representing the composition timing from the separation circuit 41 and an input line for interaction data from the interaction circuit 48 are newly added.

[0128] Next, the operation of the decoding apparatus according to the present invention will be described with reference to Figs. 18 to 21 and Fig. 48. The separation circuit 41 separates the bit stream input thereto into the compressed data and the time stamp representing the decoding timing for the audio signal, the compressed data and the time stamp representing the decoding timing for the video signal, the compressed data and the time stamp representing the decoding timing for the scene data, the time stamp representing the composition timing and the reference clock value supplied to the clock generating circuit 47 described later.

[0129] As shown in Fig. 19, in the separation circuit 41, the buffer 51 first buffers the bit stream input. Subsequently, the demultiplexer 52 separates the bit stream supplied from the buffer 51 into the compressed data and the time stamp representing the decoding timing for the audio signal, the compressed
data and the time stamp representing the decoding timing for the video signal, the compressed data and the time stamp representing the decoding timing for the scene data, the time stamp representing the composition timing, the reference clock value supplied to the clock generating circuit 47 described later and the overhead information of a system header portion on the basis of an initialization set value or control information supplied from the additive information holding circuit 53 for holding the bit stream information contained in the system header portion of the bit stream. The additive information holding circuit 53 stores the overhead information of the system header portion supplied from the demultiplexer 52 as occasion demands.

[0130] Next, the clock generating circuit 47 receives the reference clock value supplied from the separation circuit 41, and controls the oscillation frequency in accordance with the reference clock vale to generate and output clocks. However, in the case of an application for which the decoding apparatus periodically and positively fetches bit streams, for example, in such a case that the bit streams are contained in a storage medium appended to the decoding apparatus, the clock generating circuit 47 may neglect the reference clock value supplied from the separation circuit 41 and generate clocks at the oscillation frequency itself as in the case of the clock generating circuit 8.

[0131] Next, each of the audio decoding circuit 42, the video decoding circuit 43 and the scene decoding circuit 44 executes the corresponding decoding operation on the compressed data at the time given by the corresponding time stamp representing the decoding timing. As shown in Fig. 20, the decoder 63 first performs the decoding operation by using the compressed data give from the buffer 62 and the decoding data give from the memory 64 , and newly writes the decoding data thus created into the memory 64. At this time, the clocks (CLK 10) are supplied from the clock generating circuit 47 to each of the audio decoding circuit 42, the video decoding circuit 43 and the scene decoding circuit 44.

[0132] Next, the composition circuit 45 performs the composition processing at the time of the time stamp representing the composition timing supplied from the separation circuit 41 by using the audio data obtained from the audio decoding circuit 42, the video data obtained from the video decoding circuit 43 and the scene data obtained from the scene decoding circuit 43. In this case, the respective data may be directly used the decoding data stored in the memory of the decoding circuit. Further, an interaction such as movement of the viewing point for composite pictures, audio is reflected in accordance with the interaction data given from the interaction circuit 48.

[0133] The operation of Fig. 21 showing the first embodiment of the composition circuit 45 is basically the same as the circuit of Fig. 3. However, the scene generating circuit 231 starts the composition processing at the time of the time stamp representing the composition timing given from the buffer 232, and it creates scene graph by using the scene data given from the

BNSDOCID: <EP_____0924934A1_I_>

decoding circuit 44 and the interaction data given from the buffer 233 as in the case of the scene generating circuit 201, and then outputs a scene drawing command and intermediate data. The start of the operation of the other circuits can be supported by providing another control lines or setting the drawing command transmission time to the processing start time.

[0134] The operation of Fig. 48 showing the second embodiment of the composition circuit 45 is basically the same as the circuit of Fig. 45. However, CPU 22 starts the composition processing at the time of the time stamp representing the composition timing given from the separation circuit 41 through the interface circuit 49.

[0135] The operation of the display circuit 46 is the same as the display circuit 7 shown in Fig. 1. An interaction is applied to the resultingly displayed composite picture signal and audio signal through a keyboard, a mouse or the like by a viewer and the result is input to the interaction circuit 48.

[0136] Fig. 38 is a time chart showing the relationship among data of the buffer in the decoding circuit of the decoding apparatus of Fig. 18, the decoding processing on the data, data of the memory in the decoding circuit, the composition processing on the data and the final composition picture. As input compressed data are assumed first compressed video data, second compressed video data and scene data. The decoding operation on the respective data is started at the time of the time stamp representing the decoding timing. The data are read out from the buffer and the decoding processing is executed, and the decoding data thus obtained are written into the memory. Subsequently, the composition processing is started at the time of the time stamp representing the composition timing, and the respective decoding data are simultaneously read out from the memory and the composition processing is executed. The composite picture thus obtained is displayed. Fig. 39 is a time chart showing the flow of the decoding processing and the composition processing.

[0137] Fig. 39 shows a case where the processing speed of the decoding apparatus is sufficiently high and the composition is terminated within an estimated time of the coding apparatus. However when the processing speed of the decoding apparatus is not sufficient, there is a case where the composition processing needs a longer time than the estimated time of the coding apparatus. Fig. 40 is a time chart when the composition processing in the decoding apparatus needs a time above the estimated time.

[0138] As a countermeasure to the above case, the decoding and composition processing as shown in the time chart of Fig. 41 can be performed. That is, when the composition processing has not yet been terminated until the time set at the coding apparatus side, the composition is paused at the time point, that is, the time stamp representing the composition timing is neglected, and the composition is resumed at the termination time of the decoding operation. When the composition concerned has not yet been terminated until the next decoding start timing again, the composition is paused again and it is on standby until the decoding is terminated.

[0139] With respect to the audio signal and the video signal. preceding (just-before) decoding data are used for a next decoding operation, and thus skip of the decoding processing causes reduction in quality. Therefore, by pausing the composition processing as described above, the composition that causes no reduction in quality of the audio signal and the video signal can be implemented although the frame rate of the composition is reduced. However, when the pause of the composition causes missing of the audio signal in the reproduction operation, it causes great reduction in quality. Therefore, the reproduction of the audio signal in the composition is settled not to be paused.

[0140] Fig. 42 is a timing chart for the normal decoding and composition when plural input data exist, Fig. 43 is a time chart for the decoding and composition showing occurrence of the same problem as Fig. 40 when plural input data exist, and Fig. 44 is a time chart for the decoding and composition, which shows a solving method of the same problem as Fig. 41 when plural input data exist.

[0141] Fig. 22 is a block diagram showing a second embodiment of the decoding apparatus of the present invention. In this embodiment, the separation circuit 41 of Fig. 18 is replaced by a separation circuit 181, and different clock generating circuits 182, 183 and 184 are individually allocated to the decoding circuit 42 of the compressed audio data, the decoding circuit 43 for the compressed video data, and both the decoding circuit 44 for the compressed scene data and the composition 45, respectively. The separation circuit 181 is basically the same as the separation circuit 41, however, it is designed to output three reference clock values. The operation of the clock generating circuit 182, 183, 184 is the same as the clock generating circuit 47, and the oscillation frequencies thereof are controlled with the respective reference clock values given from the separation circuit 181.

[0142] As shown in Fig. 23, the separation circuit 181 is designed so that the demultiplexer 52 of Fig.19 is replaced by a demultiplexer 185. The demultiplexer 185 has three output lines for reference clock values.

[0143] Next, the operation of the circuit of Fig. 22 will be described. The basic operation is the same operation of the circuit of Fig. 18. The difference resides in that the decoding circuit 42 for the compressed audio data (hereinafter referred to as "compressed audio decoding circuit"), the decoding circuit 43 for the compressed video data (hereinafter referred to as "compressed video decoding circuit"), and both the decoding circuit 44 for the compressed scene data (hereinafter referred to as "compressed scene decoding circuit") and the composition circuit 45 are respectively operated with clocks (CLK11), (CLK12) and (CLK13) supplied

BNSDOCID: <EP_____0924934A1_I_>

from the three different clock generating circuits 182, 183 and 184, respectively, and the separation circuit 181 separates and outputs the three different reference clock values.

[0144] Fig. 24 is a block diagram showing a third embodiment of the decoding apparatus of the present invention. In this embodiment, the separation circuit 41 of Fig. 18 is replaced by the separation circuit 181. Further, the different clock generating circuits 182,183,and 184 are individually allocated to the compressed audio decoding circuit 42, both of the compressed video decoding circuit 43 and the composition circuit 45, and the compressed scene decoding circuit 44, respectively. The separation circuit 181 and the clock generating circuits 182, 183 and 184 are the same as the second embodiment of Fig. 22.

[0145] Next, the operation of the circuit of Fig. 24 will be described.

[0146] The basic operation is the same as the circuit of Fig. 18. The difference resides in that the compressed audio decoding circuit 42, both of the compressed video decoding circuit 43 and the composition circuit 45, and the compressed scene decoding circuit 44 are operated with the clocks (CLK11, CLK12, CLK13) supplied from the three different clock generating circuits 182,183 and 184, respectively, and the separation circuit 181 separates and outputs the three different reference clock values.

[0147] Fig. 25 is a block diagram showing an embodiment of the coding/decoding system in which the coding apparatus and the decoding apparatus according to the present invention are linked to each other through a transmission/storage system. In Fig. 25, the coding/decoding system comprises coding apparatus 191, decoding apparatus 192 and a transmission/storage system.

[0148] The coding apparatus 191 first receives the audio signal, the video signal and the scene data to perform the coding operation on these data, and further multiplexing the data to form a bit stream, and then transmits the multiplexed data to the transmission/storage system. Further, the decoding apparatus 192 decodes a bit stream transmitted from the transmission/storage system, receives an interaction from a viewer to perform the composition processing, and then outputs the composite picture and the audio signal.

[0149] As described above, according to the present invention, by using the time stamp representing the composition timing, a desired composite picture can be formed at the coding apparatus side and the synchronous reproduction can be performed at the decoding apparatus side. Further, when plural video signals or scene data exist and the coding/decoding is displaced in phase between these signals or data, the time stamp representing the composition timing is added to a stream of them to manage the composition timing in the decoding apparatus. Further, in accordance with complexity of the composition, the decoding operation and the composition operation of the decoding apparatus can be controlled at the coding apparatus side.

[0150] It is unnecessary to provide the two time stamps of the time stamp representing the decoding timing and the time stamp representing the composition timing, and by using one flag it may be informed whether the stream concerned is a stream for managing the composition processing or not. As described above, use of the flag can avoid necessity of inserting the time stamp representing the composition timing into the bit stream, and thus the bit amount can be reduced. In this case, it is assumed that the decoding timing and the composition timing are coincident with each other.

[0151] Fig. 49 shows an embodiment of the bit stream of the present invention when the 1-bit flag as described above is used. a 1-bit flag is added to the time stamp representing the decoding timing which is appended to each of the compressed audio data, the compressed video data and the compressed scene data, and then the multiplexing operation is carried out to generate a bit stream.

[0152] It is assumed that when the flag is "0", it is assumed that the time stamp representing the decoding timing does not double as the time stamp representing the composition timing while when the flag is "1", the time stamp representing the decoding timing doubles as the time stamp representing the composition timing.

[0153] Fig. 50 shows another embodiment of the bit stream according to the present invention in which the 1-bit flag is added to the reference clock value and the time stamp representing the decoding timing. The 1-bit flag is added to the reference clock value and the time stamp representing the decoding timing which is appended to each of the compressed audio data, the compressed video data and the compressed scene data, and the multiplexing operation is carried out to generate a bit stream.

[0154] It is assumed that when the flag is "0", the time stamp representing the decoding timing does not double as the time stamp representing the composition timing while when the flag is "1", the time stamp representing the decoding timing doubles as the time stamp representing the composition timing.

[0155] Fig. 51 is a block diagram showing a tenth embodiment of the coding apparatus according to the present invention.

[0156] According to this embodiment, in the construction of Fig. 5, the video coding circuit 151, the scene coding circuit 4, the composition circuit 152 and the multiplexing circuit 6 are replaced by a coding circuit 241, a coding circuit 242, a composition circuit 243 and a multiplexing circuit 244.

[0157] Next, the operation of the circuit of Fig. 51 will be described.

[0158] The operation of the circuit of Fig. 51 is basically the same as that of Fig. 5. However, the video coding circuit 241 and the scene coding circuit 242 set the flag of the bit stream of the present invention to "1" and

BNSDOCID: <EP_____0924934A1_I_>

outputs it as time information together with the time stamps representing the decoding timing when the streams thereof carry the composition timing. Conversely, when the streams do not carry the composition timing, the flag of the bit stream of the present invention is set to "0", and output as time information together with the time stamp representing the decoding timing. The composition circuit 243 outputs the composition status as in the case of the composition circuit 152 of Fig. 5. On the other hand, when the composition processing of the composition circuit 243 is not terminated, the video coding circuit 241 or the scene coding circuit 242 sets the flag of the bit stream of the present invention to "0" and outputs it as time information together with the time stamp representing the decoding timing even if the stream originally carries the composition timing. The multiplexing circuit 244 generates and outputs the bit stream according to the present invention.

[0159] Fig. 52 is a block diagram showing a fourth embodiment of the decoding apparatus according to the present invention. In this embodiment, in the construction of Fig. 18, the separation circuit 41 is replaced by a separation circuit 251. The separation circuit 251 copies and outputs the time stamp representing the decoding timing of a stream which carries the composition timing.

[0160] Next, the operation of the circuit of Fig. 52 will be described.

[0161] The operation of the circuit of Fig. 52 is basically the same as Fig. 18. However, according to the flag of the bit stream of the presents invention, the separation circuit 251 copies and outputs the time stamp representing the decoding timing of a stream which carries the composition timing. The composition circuit 45 starts the composition operation in accordance with the time stamp. However, actually, it waits until the termination of the processing of the decoding circuit which decodes the stream carrying the composition timing, and starts the composition processing just after the termination of the processing.

[0162] Further, the coding apparatus and the decoding apparatus shown in Figs. 51 and 52 may be linked to each other to fabricate the coding/decoding system shown in Fig. 25.

[0163] According to the coding apparatus of the present invention, the time stamp representing the composition timing is added to the bit stream. Therefore, the generation of a desired composition picture at the coding side can be ensured, and the stream data that are transmitted continuously on time axis can be supported. In addition, the coding /decoding synchronous reproduction of audio signals, video signals and artificial scene data can be implemented with supporting the interaction function at the decoding side.

[0164] According to the second embodiment of the coding apparatus of the present invention, when the composition load is high, the coding processing of the video signal is controlled and the time stamp representing the composition is added to the bit stream. Therefore, the generation of a composite picture desired at the coding side can be ensured and the stream data that are transmitted continuously on time axis can be supported. In addition, the coding/decoding synchronous reproduction of audio signals, video signals and artificial scene data can be implemented with supporting the interaction function at the decoding side and without reducing the composition frame rate.

[0165] According to the third embodiment of the coding apparatus of the present invention, when the composition load is high, the composition processing is controlled, and the time stamp representing the composition timing is added to the bit stream. Therefore, the generation of a composite picture desired at the coding side can be ensured and the stream data that are transmitted continuously on time axis can be supported. In addition, the coding/decoding synchronous reproduction of audio signals, video signals and artificial scene data can be implemented with supporting the interaction function at the decoding side and without reducing the frame rate of video signal.

[0166] According to the fourth embodiment of the coding apparatus of the present invention, the same clocks are supplied to the composition circuit and the coding circuit for artificial scene data, and the time stamp representing the composition timing is added to the compressed data of the artificial scene data to generate a bit stream. Therefore, the generation of a composite picture desired at the coding side can be ensured and the stream data that are transmitted continuously on time axis can be supported. In addition, the coding/decoding synchronous reproduction of audio signals, video signals and artificial scene data when the coding is performed with clocks which are different among the audio signal, the video signal and the artificial scene data can be implemented with supporting the interaction function at the decoding side.

[0167] According to the fifth embodiment of the coding apparatus of the present invention, when the composition load is high, the coding processing of the video signal is controlled, the same clocks are supplied to the composition circuit and the coding circuit for the artificial scene data, and the time stamp representing the composition timing is appended to the compressed data of the artificial scene data to generate a bit stream. Therefore, the generation of a composite picture desired at the coding side can be ensured and the stream data that are transmitted continuously on time axis can be supported. In addition, the coding/decoding synchronous reproduction of audio signals, video signals and artificial scene data when the coding is performed with clocks which are different among the audio signal, the video signal and the artificial scene data can be implemented with supporting the interaction function at the decoding side and without reducing the composition frame rate.

[0168] According to the sixth embodiment of the cod-

BNSDOCID: <EP_____0924934A1_I_>

ing apparatus of the present invention, when the composition load is high, the composition processing is controlled, the same clocks are supplied to the composition circuit and the coding circuit for the artificial scene data, and the time stamp representing the composition timing is appended to the compressed data of the artificial scene data to generate a bit stream. Therefore, the generation of a composite picture desired at the coding side can be ensured and the stream data which are transmitted continuously on time axis can be supported. In addition, the coding/decoding synchronous reproduction of audio signals, video signals and artificial scene data when the coding is performed with clocks which are different among the audio signal, the video signal and the artificial scene data can be implemented with supporting the interaction function at the decoding side and without reducing the frame rate of video signal.

[0169] According to the seventh embodiment of the coding apparatus of the present invention, the same clocks are supplied to the composition circuit and the coding circuit for the video signal, and the time stamp representing the composition timing is appended to the compressed data of the video signal to generate a bit stream. Therefore, the generation of a composite picture desired at the coding side can be ensured and the stream data that are transmitted continuously on time axis can be supported. In addition, the coding/decoding synchronous reproduction of audio signals, video signals and artificial scene data when the coding is performed with clocks which are different among the audio signal, the video signal and the artificial scene data can be implemented with supporting the interaction function at the decoding side.

[0170] According to the eighth embodiment of the coding apparatus of the present invention, when the composition load is high, the coding processing of the video signal is controlled, the same clocks are supplied to the composition circuit and the coding circuit for the video signal, and the time stamp representing the composition timing is appended to the compressed data of the video signal to generate a bit stream. Therefore, the generation of a composite picture desired at the coding side can be ensured and the stream data which are transmitted continuously on time axis can be supported. In addition, the coding/decoding synchronous reproduction of audio signals, video signals and artificial scene data when the coding is performed with clocks which are different among the audio signal, the video signal and the artificial scene data can be implemented with supporting the interaction function at the decoding side and without reducing the composition frame rate.

[0171] According to the ninth embodiment of the coding apparatus of the present invention, when the composition load is high, the composition processing is controlled, the same clocks are supplied to the composition circuit and the coding circuit for the video signal, and the time stamp representing the composition timing is appended to the compressed data of the video signal to generate a bit stream. Therefore, the generation of a composite picture desired at the coding side can be ensured and the stream data which are transmitted continuously on time axis can be supported. In addition, the coding/decoding synchronous reproduction of audio signals, video signals and artificial scene data when the coding is performed with clocks which are different among the audio signal, the video signal and the artificial scene data can be implemented with supporting the interaction function at the decoding side and without reducing the frame rate of video signal.

[0172] According to the decoding apparatus of the present invention, the composition processing is performed by using the time stamp representing the composition timing that is added to the bit stream. Therefore, the generation of a composite picture desired at the coding side can be ensured and the stream data that are transmitted continuously on time axis can be supported. In addition, the coding/decoding synchronous reproduction of audio signals, video signals and artificial scene data can be implemented with supporting the interaction function at the decoding side.

[0173] According to the second embodiment of the decoding apparatus of the present invention, the composition circuit and the decoding apparatus for the compressed artificial scene data are driven by using clocks generated with a reference clock value which is appended to the compressed data of the artificial scene data in the bit stream, and the composition processing is performed by using the time stamp representing the composition timing appended to the compressed data of the artificial scene data. Therefore, the generation of a composite picture desired at the coding side can be ensured and the stream data that are transmitted continuously on time axis can be supported. In addition, the coding/decoding synchronous reproduction of audio signals, video signals and artificial scene data when the coding is performed with clocks which are different among the audio signal, the video signal and the artificial scene data can be implemented with supporting the interaction function at the decoding side.

[0174] According to the third embodiment of the decoding apparatus of the present invention, the composition circuit and the decoding apparatus for the compressed data of the video signal are driven by using clocks generated with a reference clock value which is appended to the compressed data of the video signal in the bit stream, and the composition processing is performed by using the time stamp representing the composition timing appended to the compressed data of the video signal. Therefore, the generation of a composite picture desired at the coding side can be ensured and the stream data that are transmitted continuously on time axis can be supported. In addition, the coding/decoding synchronous reproduction of audio signals, video signals and artificial scene data when the coding is performed with clocks which are different among the audio signal, the video signal and the artifi-

BNSDOCID: <EP_____0924934A1_I_>

cial scene data can be implemented with supporting the interaction function at the decoding side.

[0175] According to the coding/decoding system of the present invention, the coding/decoding system is constituted by proper combination of the coding apparatus of the present invention and the decoding apparatus of the present invention. Therefore, the generation of a composite picture desired at the coding side can be ensured and the stream data that are transmitted continuously on time axis can be supported. In addition, the coding/decoding synchronous reproduction of audio signals, video signals and artificial scene data can be implemented with the operation/working-effect by the combination of the coding apparatus and the decoding apparatus and with supporting the interaction function at the decoding side.

[0176] According to the bit stream of the present invention, the time stamp representing the decoding timing and the time stamp representing the composition timing can be made common to each other. Therefore, the generation of a composite picture desired at the coding side can be ensured and the stream data that are transmitted continuously on time axis can be supported. In addition, the coding/decoding synchronous reproduction of audio signals, video signals and artificial scene data when the coding is performed with clocks which are different among the audio signal, the video signal and the artificial scene data can be implemented with supporting the interaction function at the decoding side and reducing overhead information.

[0177] According to the tenth embodiment of the coding apparatus of the present invention, the time stamp representing the decoding timing and the time stamp representing the composition timing are made common by using a flag to generate a bit stream. Therefore, the generation of a composite picture desired at the coding side can be ensured and the stream data that are transmitted continuously on time axis can be supported. In addition. the coding/decoding synchronous reproduction of audio signals, video signals and artificial scene data when the coding is performed with clocks which are different among the audio signal, the video signal and the artificial scene data can be implemented with supporting the interaction function at the decoding side and reducing overhead information.

[0178] According to the fourth embodiment of the decoding apparatus of the present invention, the decoding processing is performed by using the bit stream which is obtained by making common the time stamp representing the decoding timing and the time stamp representing the composition timing with a flag. Therefore, the generation of a composite picture desired at the coding side can be ensured and the stream data that are transmitted continuously on time axis can be supported. In addition, the coding/decoding synchronous reproduction of audio signals, video signals and artificial scene data when the coding is performed with clocks which are different among the audio signal, the video signal and the artificial scene data can be implemented with supporting the interaction function at the decoding side and reducing overhead information.

[0179] According to the another embodiment of the coding/decoding system of the present invention, it uses the coding apparatus and the decoding apparatus using the bit stream which is obtained by making common the time stamp representing the decoding timing and the time stamp representing the composition timing with a flag. Therefore, the generation of a composite picture desired at the coding side can be ensured and the stream data that are transmitted continuously on time axis can be supported. In addition, the coding/decoding synchronous reproduction of audio signals, video signals and artificial scene data when the coding is performed with clocks which are different among the audio signal, the video signal and the artificial scene data can be implemented with supporting the interaction function at the decoding side and reducing overhead information.

Claims

1. A coding apparatus comprising:

   audio signal coding means for coding an audio signal;
   video signal coding means for coding a video signal;
   interface means for accepting information on a composite scene;
   scene data coding means for coding scene data supplied from said interface means;
   composition means for composing a scene from the audio signal supplied from said audio signal coding means, the video signal supplied from said video signal coding means and the composite scene data supplied from said scene data coding means;
   display means for reproducing/displaying the composite picture signal and the audio signal supplied from said composition means;
   clock supply means for supplying clocks to said audio signal coding means, said video signal coding means, said scene data coding means and said composition means; and
   multiplexing means for creating a bit stream on the basis of the time information and compressed audio data supplied from said audio signal coding means, the time information and compressed video data supplied from said video signal coding means, the time information and compressed scene data supplied from said scene data coding means, the time information supplied from said composition means and the clock value supplied from said clock supplying means.

BNSDOCID: <EP_____0924934A1_I_>

2. The coding apparatus as claimed in claim 1, further comprising means for detecting the status of said composition means and controlling the operation of said video signal coding means.

3. The apparatus as claimed in claim 1 or 2, further comprising means for detecting the status of said audio signal coding means, the status of said video signal coding means and the status of said scene data coding means, and controlling the operation of said composition means.

4. The apparatus as claimed in claim 1, 2 or 3, wherein said clock supply means includes first clock supply means for supplying clocks to said audio signal coding means, second clock supply means for supplying clocks to said video signal coding means and third clock supply means for supplying clocks to said scene data coding means and composition means, and said multiplexing means multiplexes the clock values supplied from said first, second, and third clock supply means respectively.

5. The apparatus as claimed in claim 1, 2 or 3, wherein said clock supply means includes first clock supply means for supplying clocks to said audio signal coding means, second clock supply means for supplying clocks to said video signal coding means and composition means, and third clock supply means for supplying clocks to said scene data coding means, and said multiplexing means multiplexes the clock values supplied from said first, second, and third clock supply means respectively.

6. A decoding apparatus comprising:

means for separating both of compressed data and time information of an audio signal, both of compressed data and time information of a video signal, both of compressed data and time information of scene data, time information of scene composition and clock information from a bit stream;
means for decoding the audio signal on the basis of the compressed data and time information of said audio signal;
means for decoding the video signal on the basis of the compressed data and time information of the video signal;
means for decoding the scene data on the basis of the compressed data and time information of the scene data;
means for composing a scene on the basis of the time information for the scene composition supplied from said separation means, the audio signal supplied from said decoding means for the audio signal, the video signal supplied from said decoding means for the video signal and the scene data supplied from said decoding means for the scene data;
means for generating clocks according to the clock value supplied from said separating means and supplying the clocks to said decoding means for the audio signal, said decoding means for the video signal, said decoding means for the scene data and said composition means;
means for reproducing/displaying the composite picture signal and the audio signal supplied from said composition means; and
interface means for accepting an interaction from a viewer to the composite picture.

7. The decoding apparatus as claimed in claim 6, wherein said separation means separates a plurality of independent clock values from said bit stream, and the independent clock values are input to means for supplying the clocks to said decoding means for the audio signal, means for supplying the clocks to said decoding means for the video signal, and means for supplying the clocks to said decoding means for the scene data and said composition means.

8. The decoding apparatus as claimed in claim 6, wherein said separation means separates a plurality of independent clock values from said bit stream, and the independent clock values are input to means for supplying the clocks to said decoding means for the audio signal, means for supplying the clocks to said decoding means for the video signal and said composition means, and means for supplying the clocks to said decoding means for the scene data.

9. A coding/decoding system comprising said coding apparatus as claimed in any one of claims 1 to 5 and said decoding apparatus as claimed in claim 6, 7 or 8.

10. A multiplexed bit stream comprising an audio signal, a video signal and scene data, characterized in that a flag representing whether time information representing a decoding timing doubles as time information representing a composition timing is added to said time information.

11. The coding apparatus as claimed in any one of claims 1 to 5, wherein said coding apparatus generates said bit stream as claimed in claim 9.

12. The decoding apparatus as claimed in claim 6, 7 or 8, wherein said decoding apparatus decodes said bit stream as claimed in claim 9.

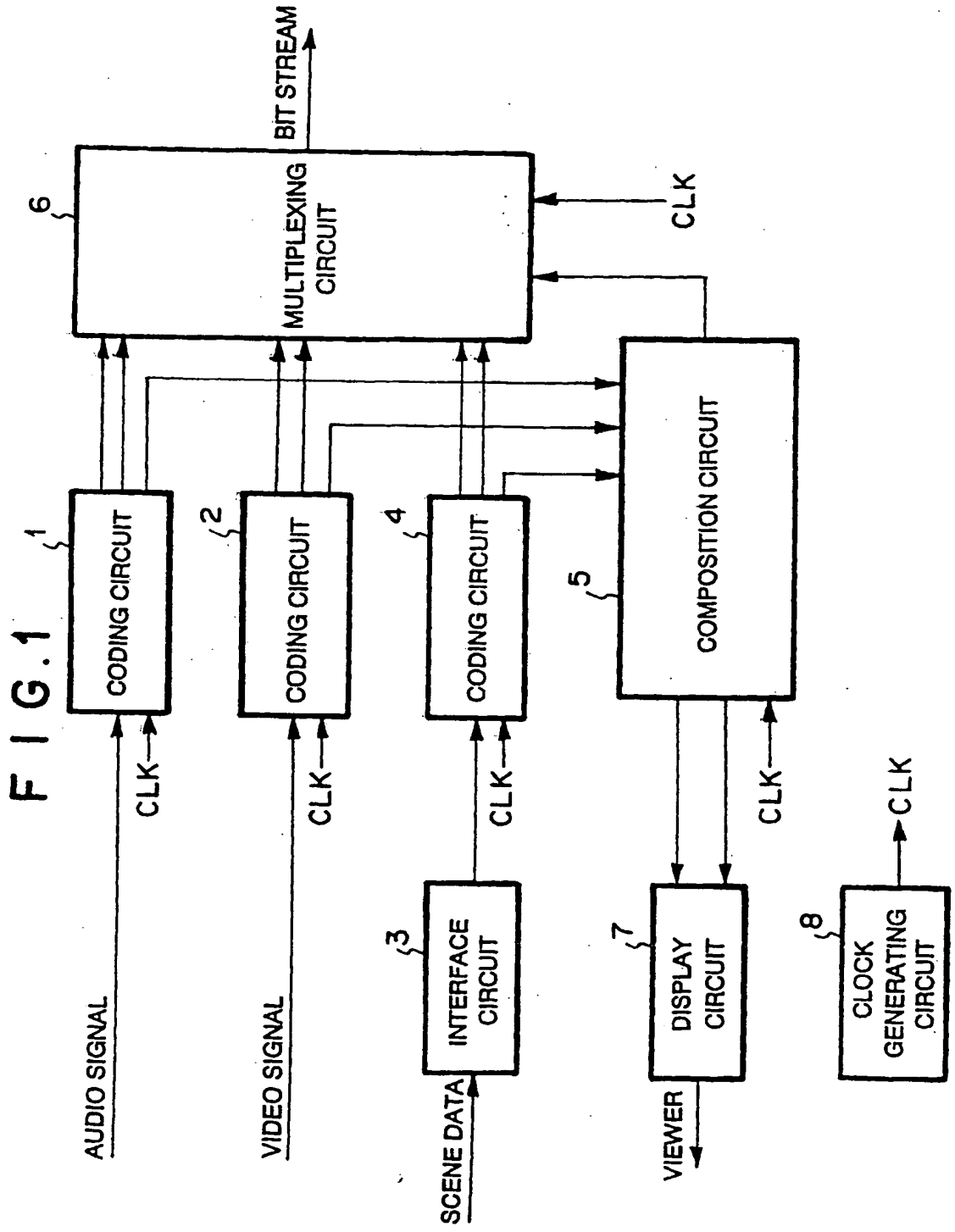
13. A coding/decoding system comprising said coding

BNSDOCID: <EP_____0924934A1_I_>

apparatus as claimed in claim 11 and said decoding apparatus as claimed in claim 12.

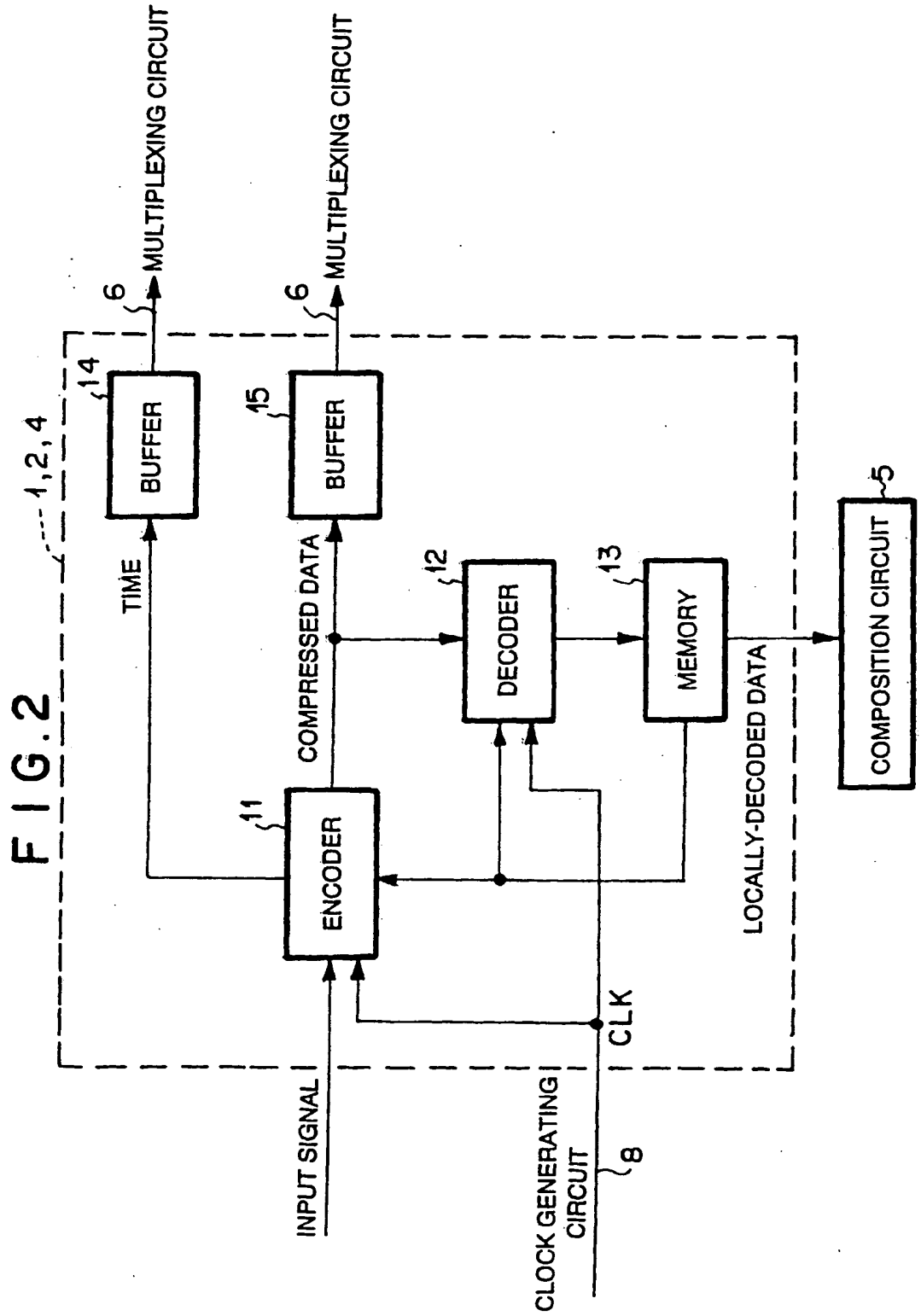
BNSDOCID: <EP_____0924934A1_I_>

FIG.1



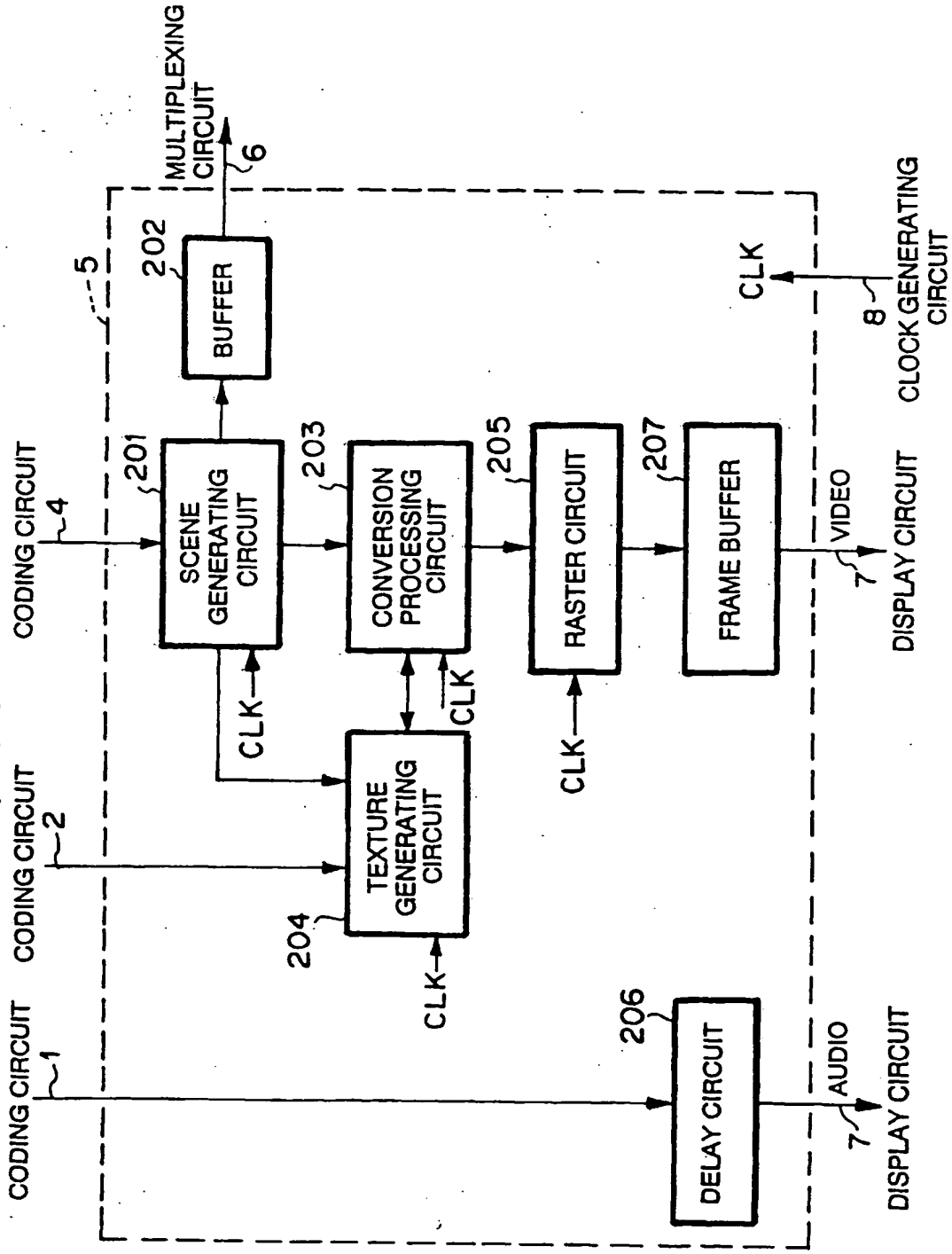
BNSDOCID: <EP_____0924934A1_I_>

FIG.2



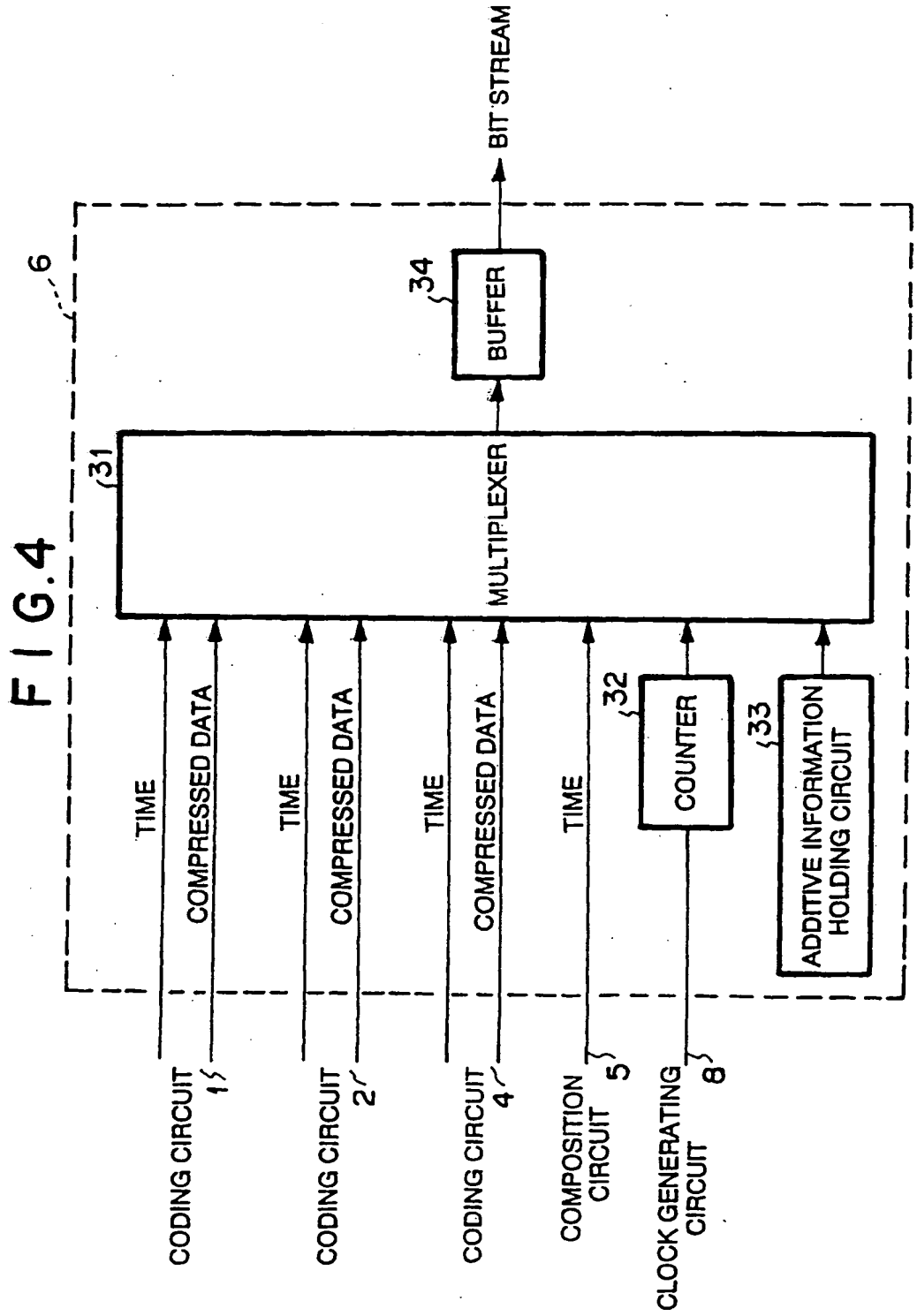
BNSDOCID: <EP_____0924934A1_I_>

FIG. 3



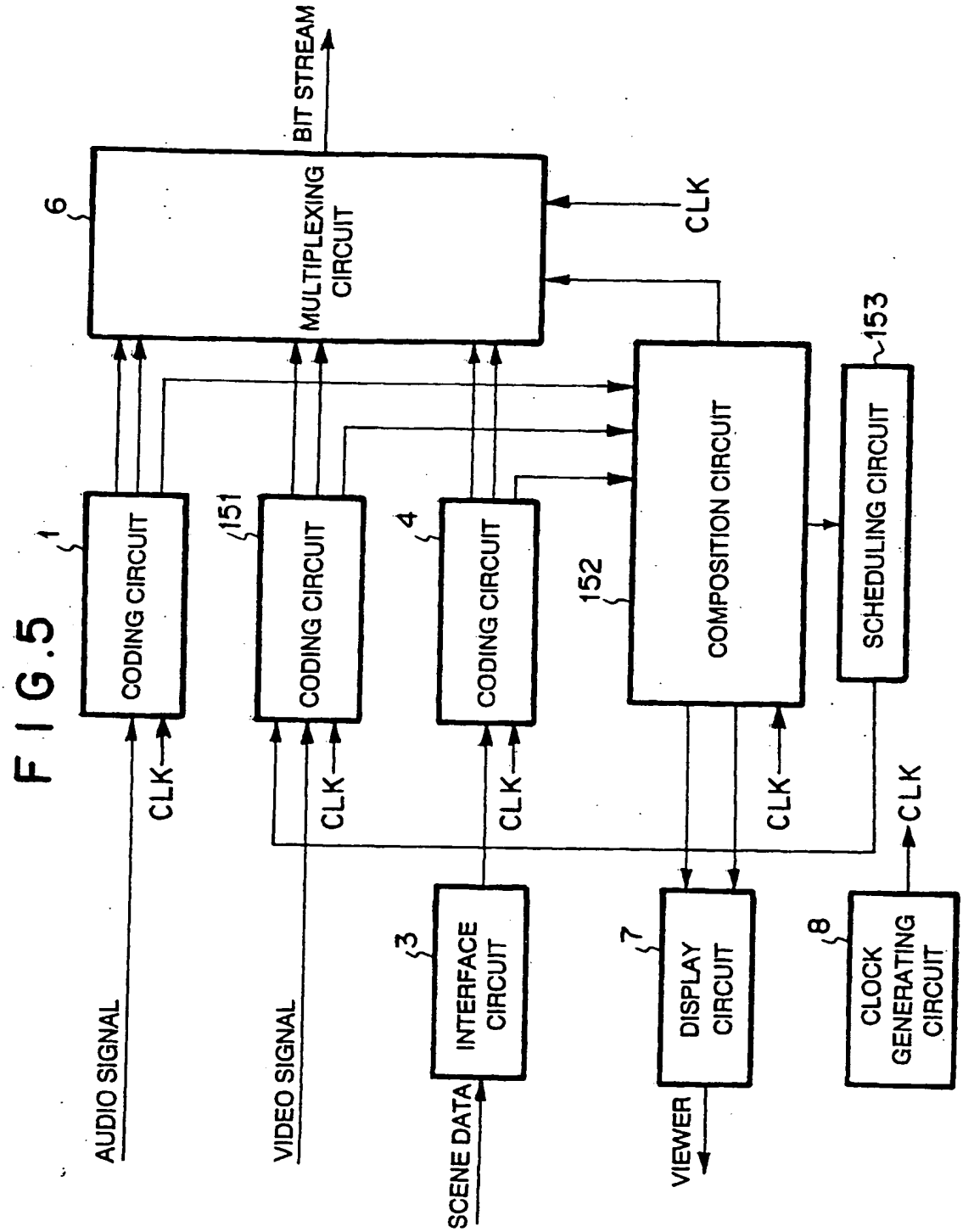
BNSDOCID: <EP_____0924934A1_I_>

FIG.4



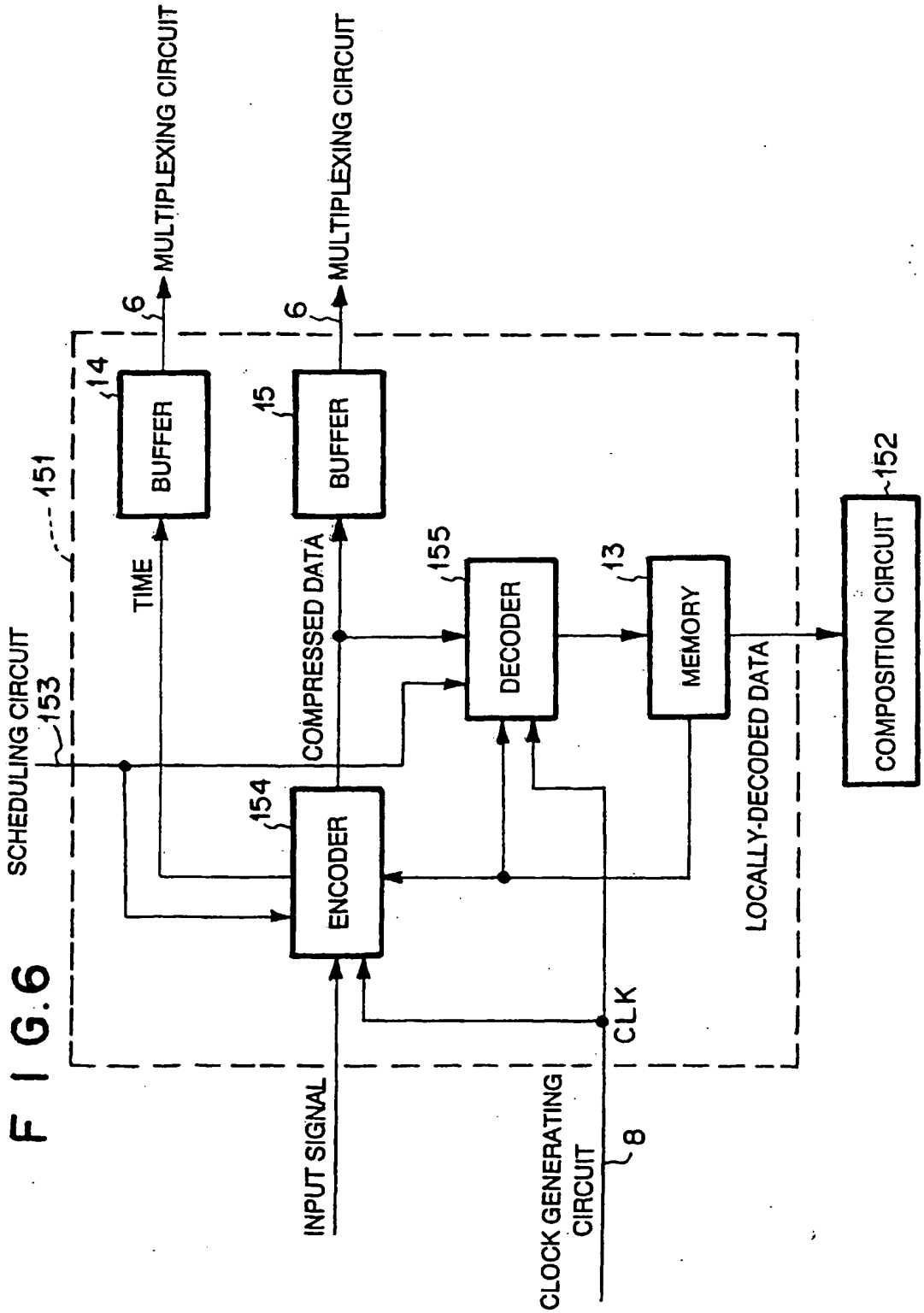
BNSDOCID: <EP_____0924934A1_I_>

FIG.5



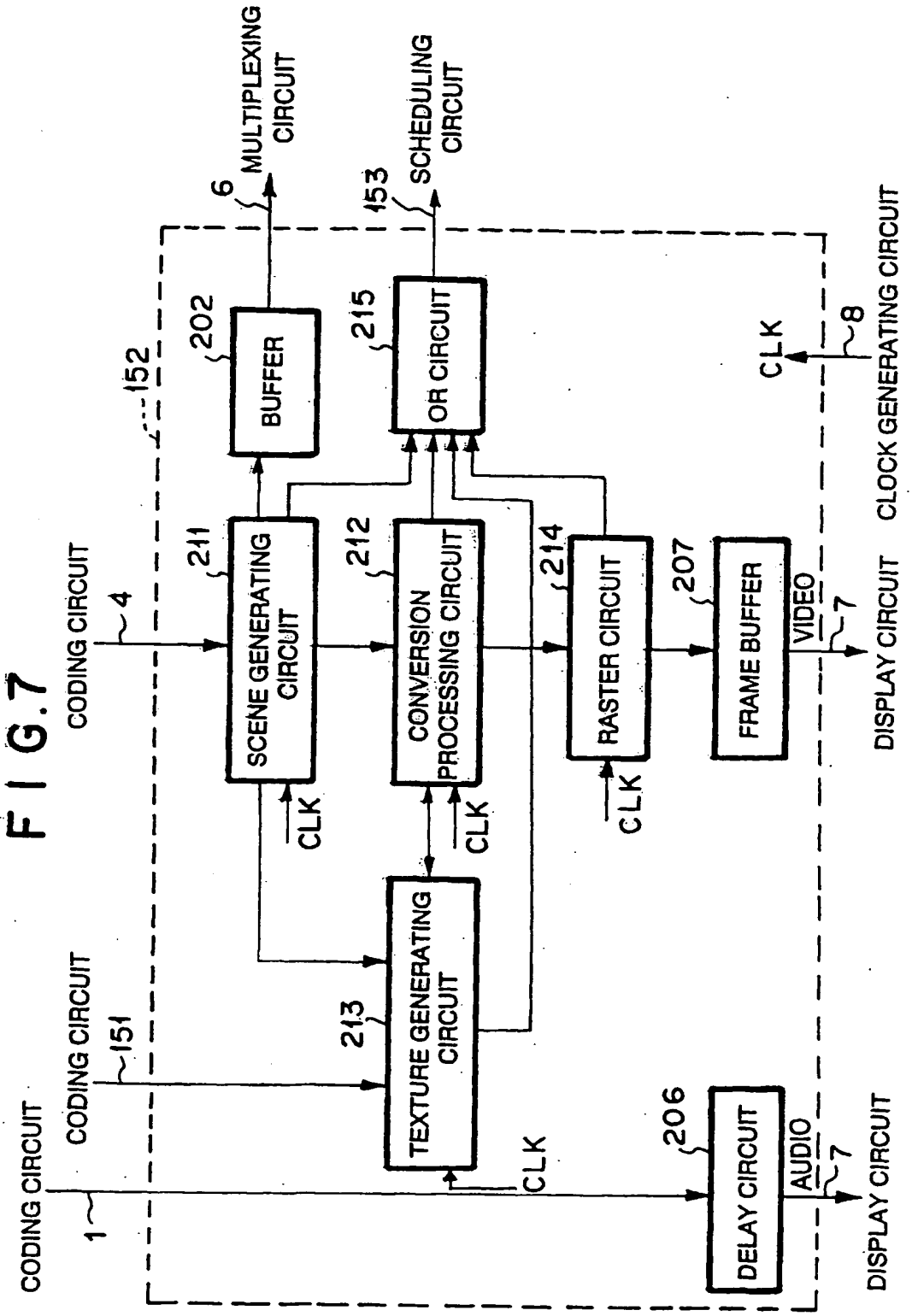
BNSDOCID: <EP_____0924934A1_I_>

FIG.6



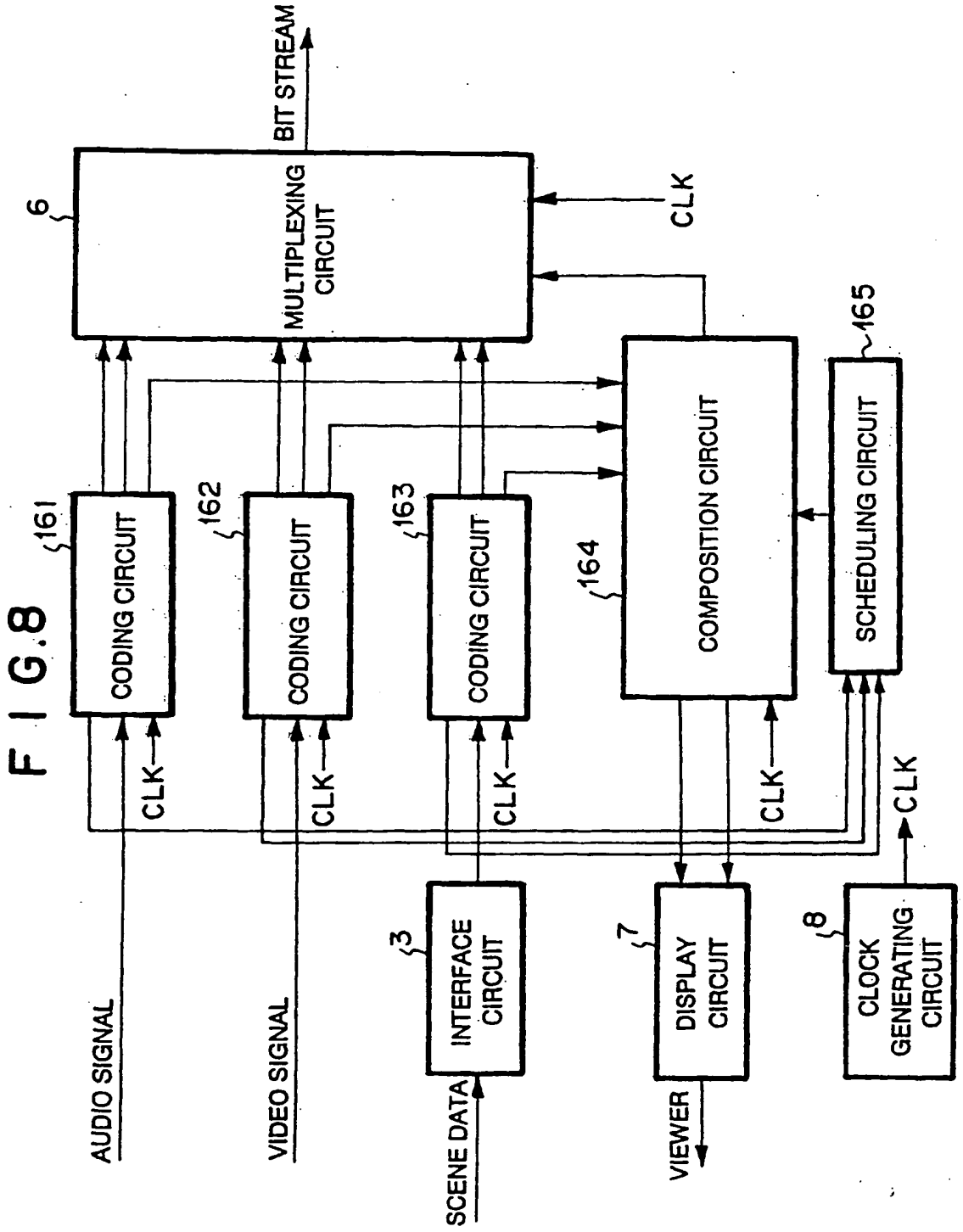
BNSDOCID: <EP_____0924934A1_I_>

FIG.7



BNSDOCID: <EP_____0924934A1_I_>

FIG.8



BNSDOCID: <EP_____0924934A1_I_>

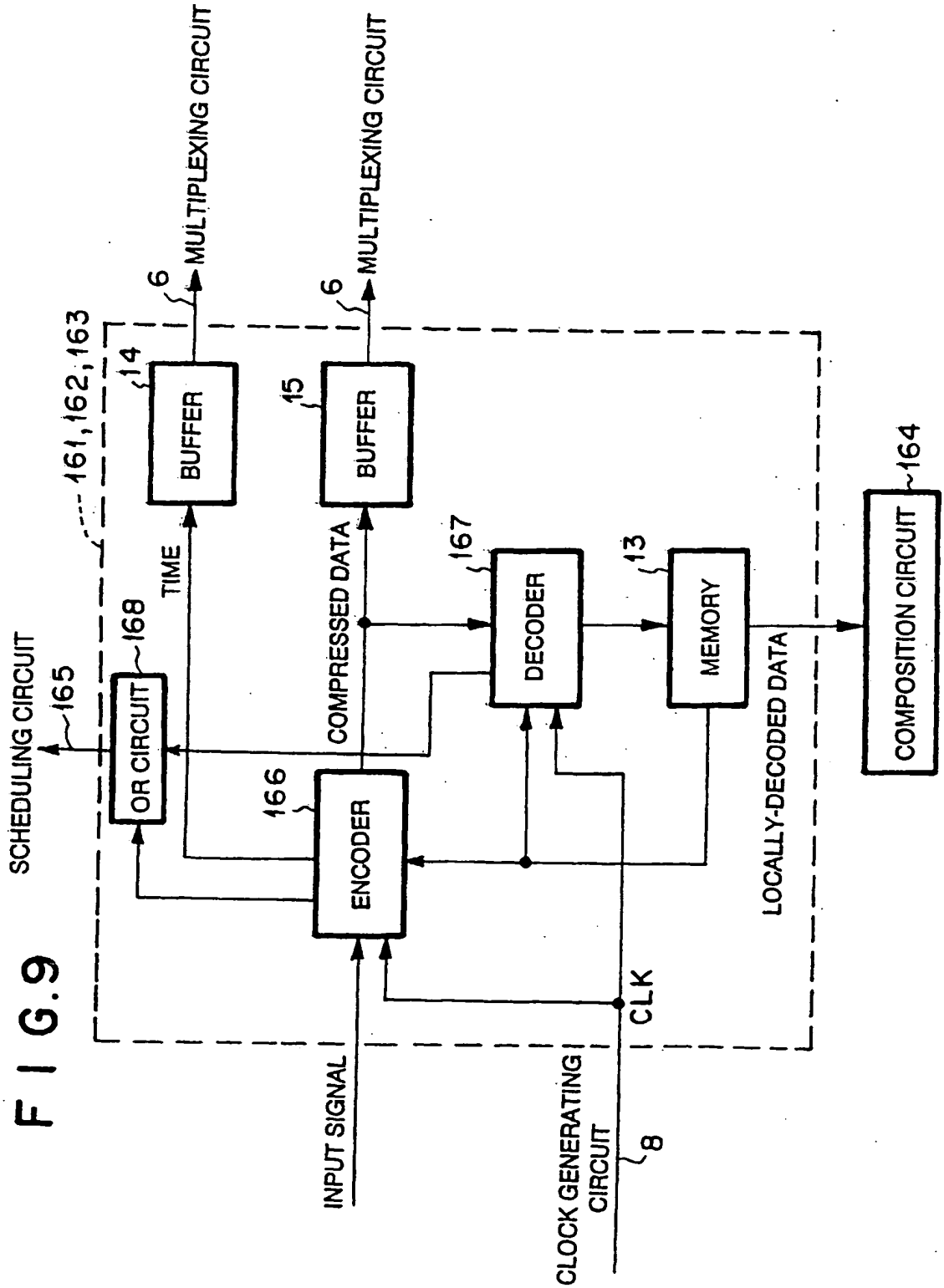
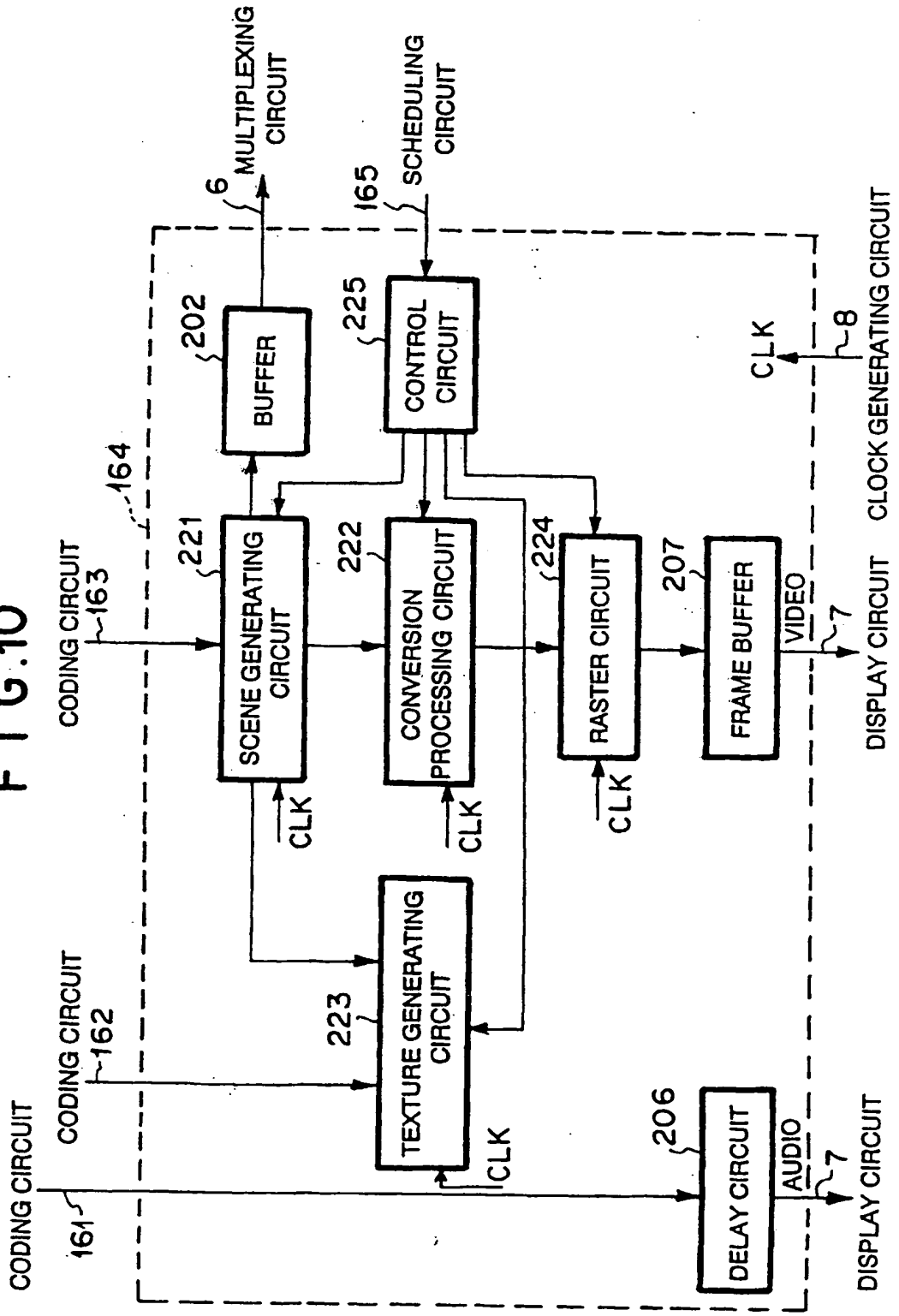


FIG.9

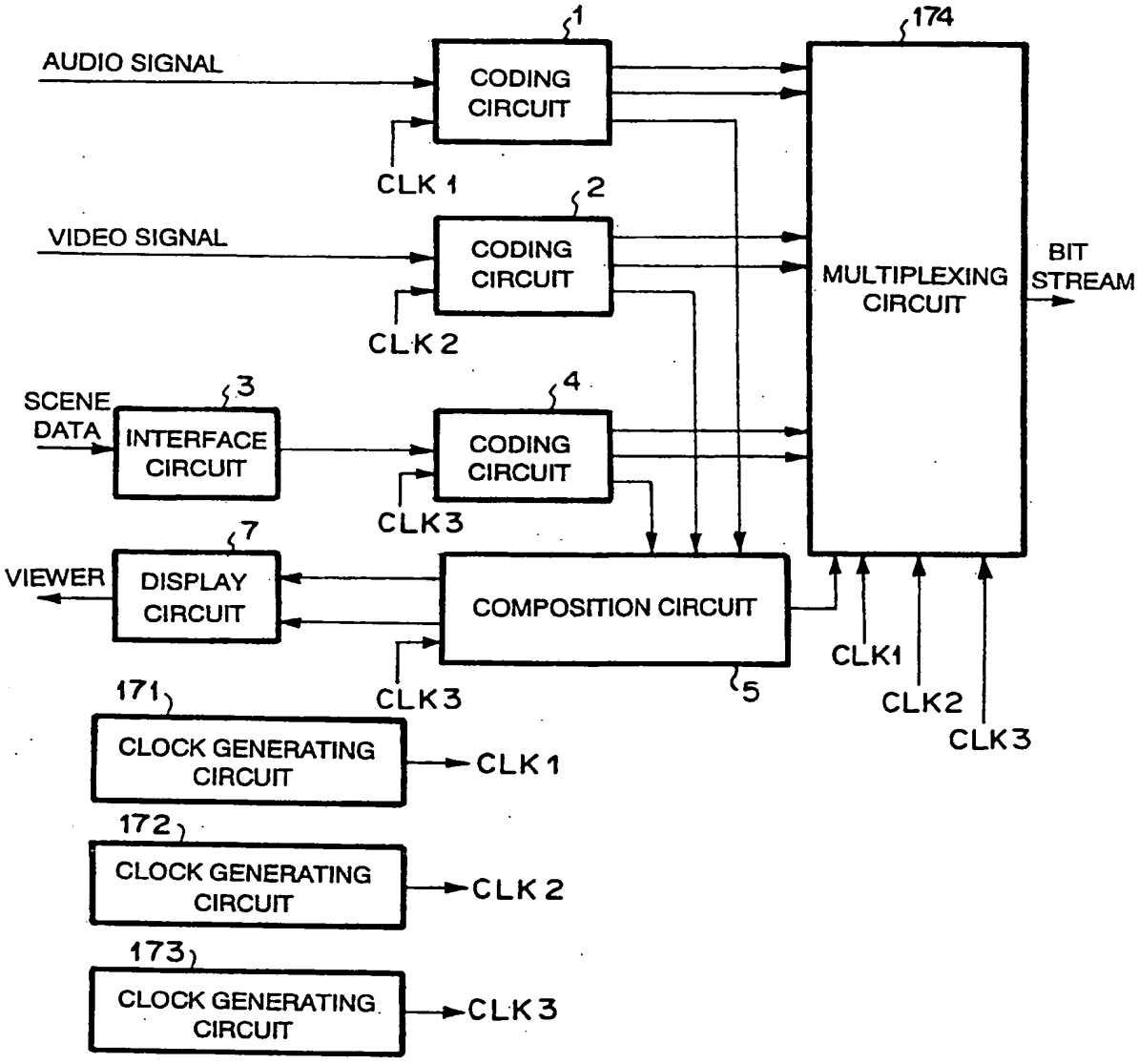
BNSDOCID: <EP_____0924934A1_I_>

FIG.10



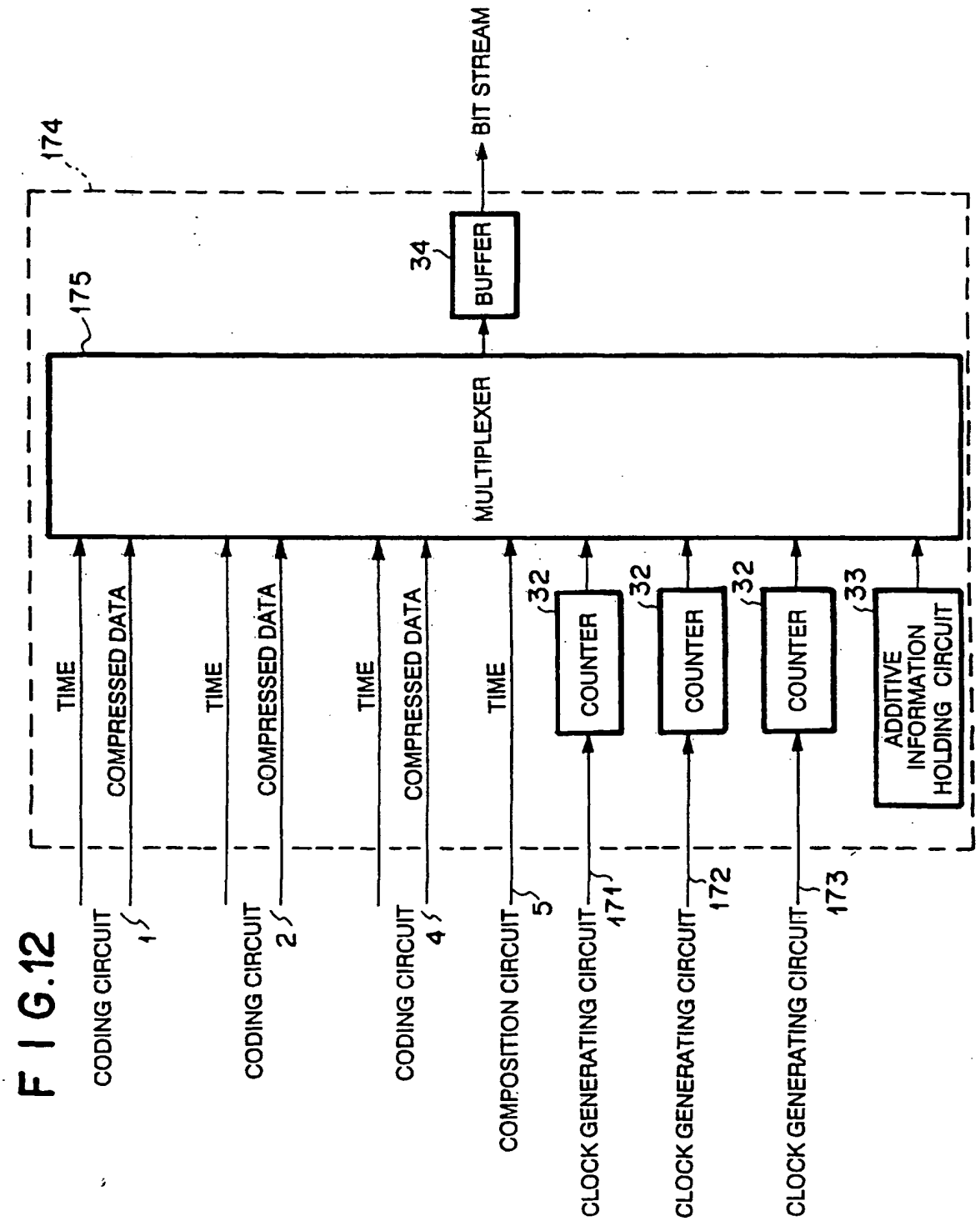
BNSDOCID: <EP_____0924934A1_I_>

FIG.11



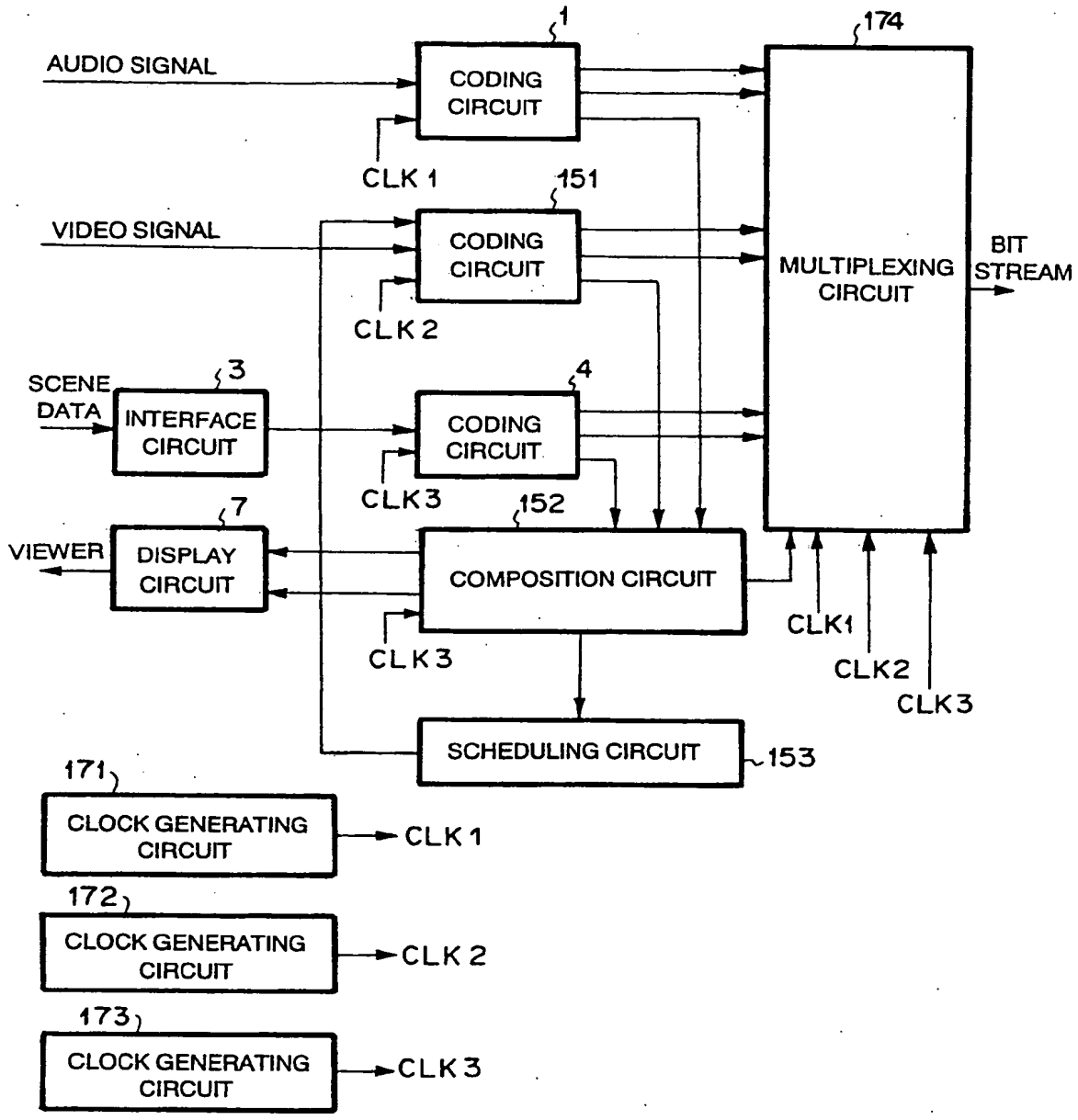
3NSDOCID: <EP_____0924934A1_I_>

FIG.12



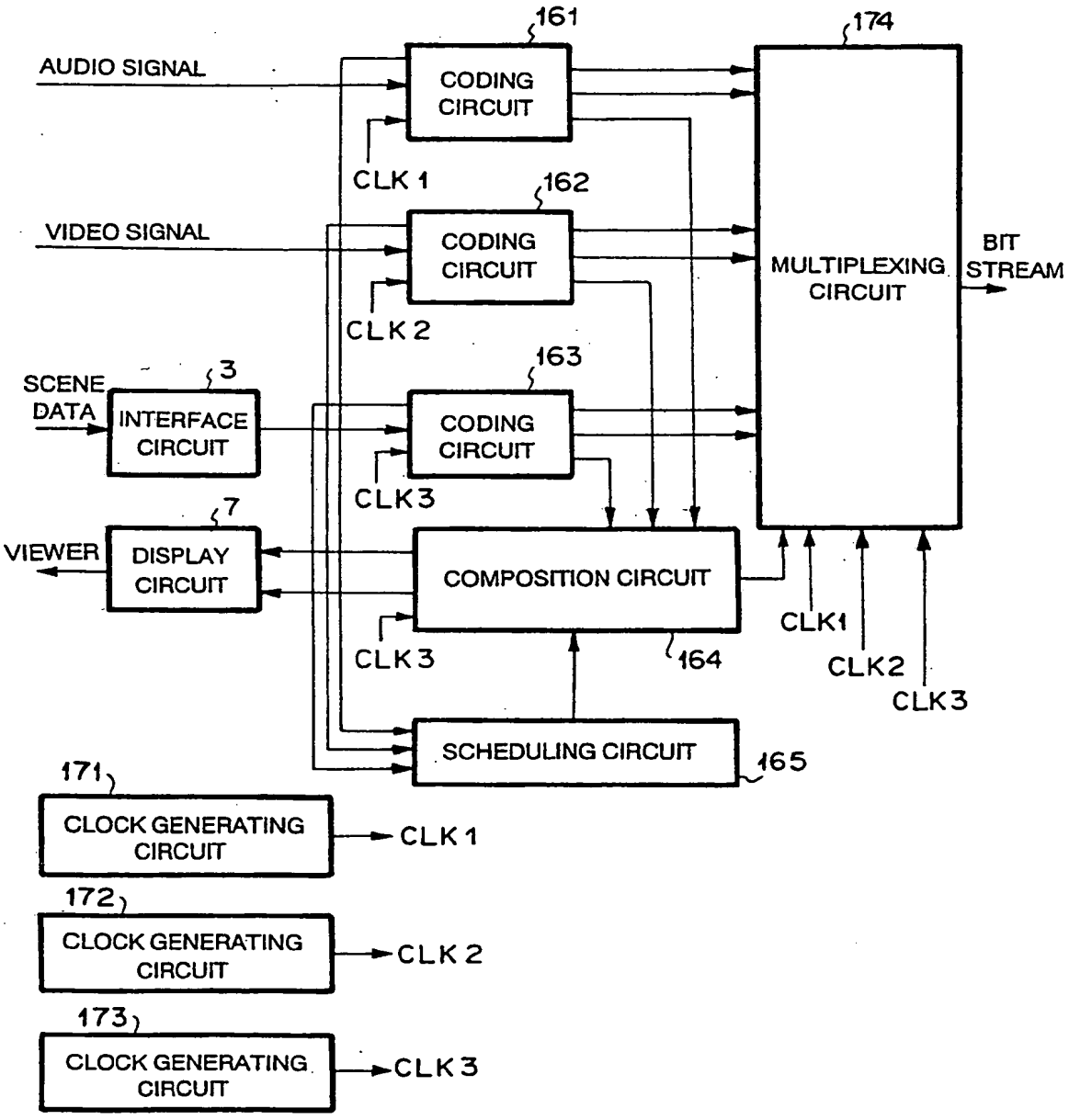
BNSDOCID: <EP_____0924934A1_I_>

FIG.13



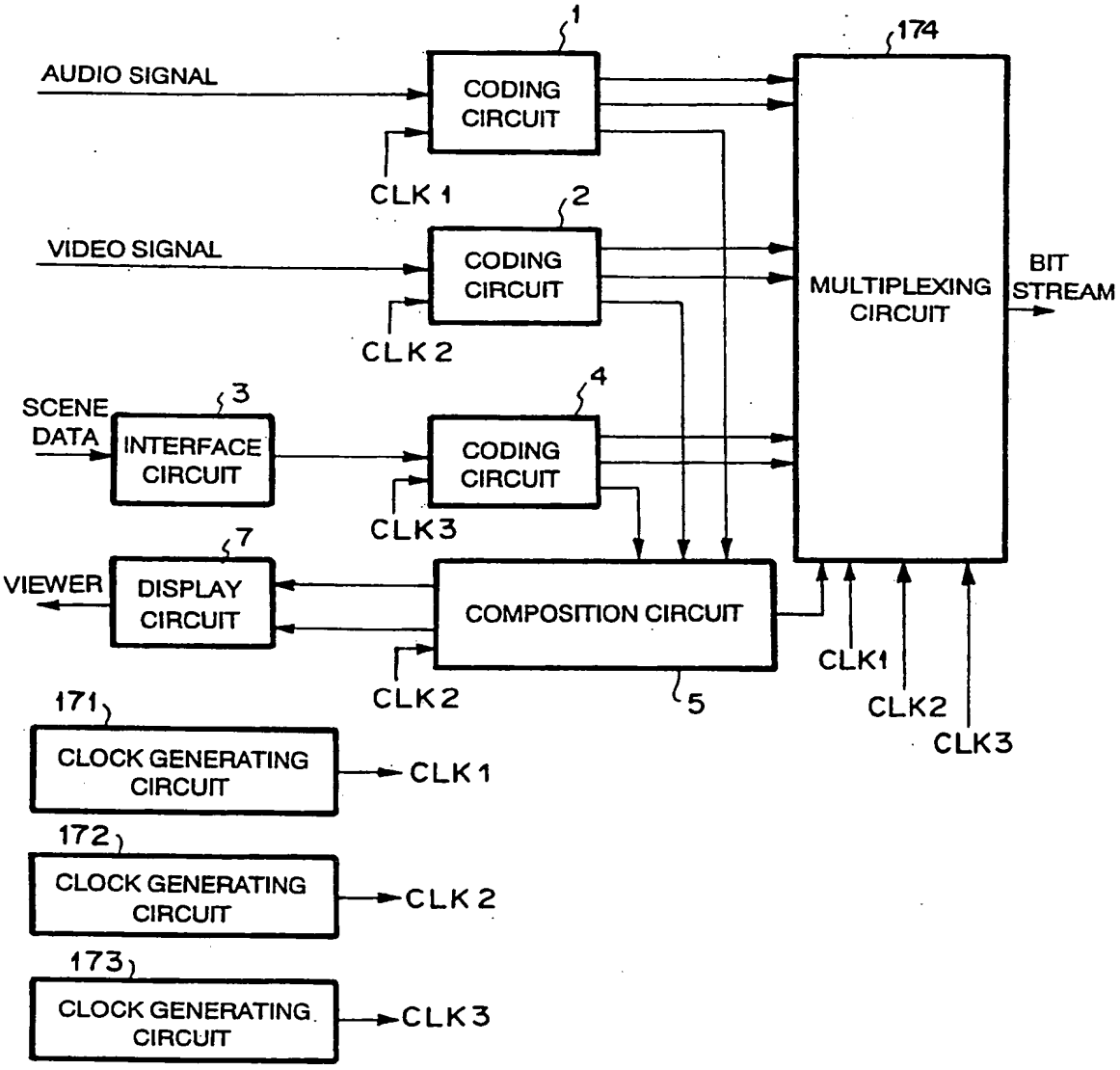
BNSDOCID: <EP_____0924934A1_I_>

FIG.14



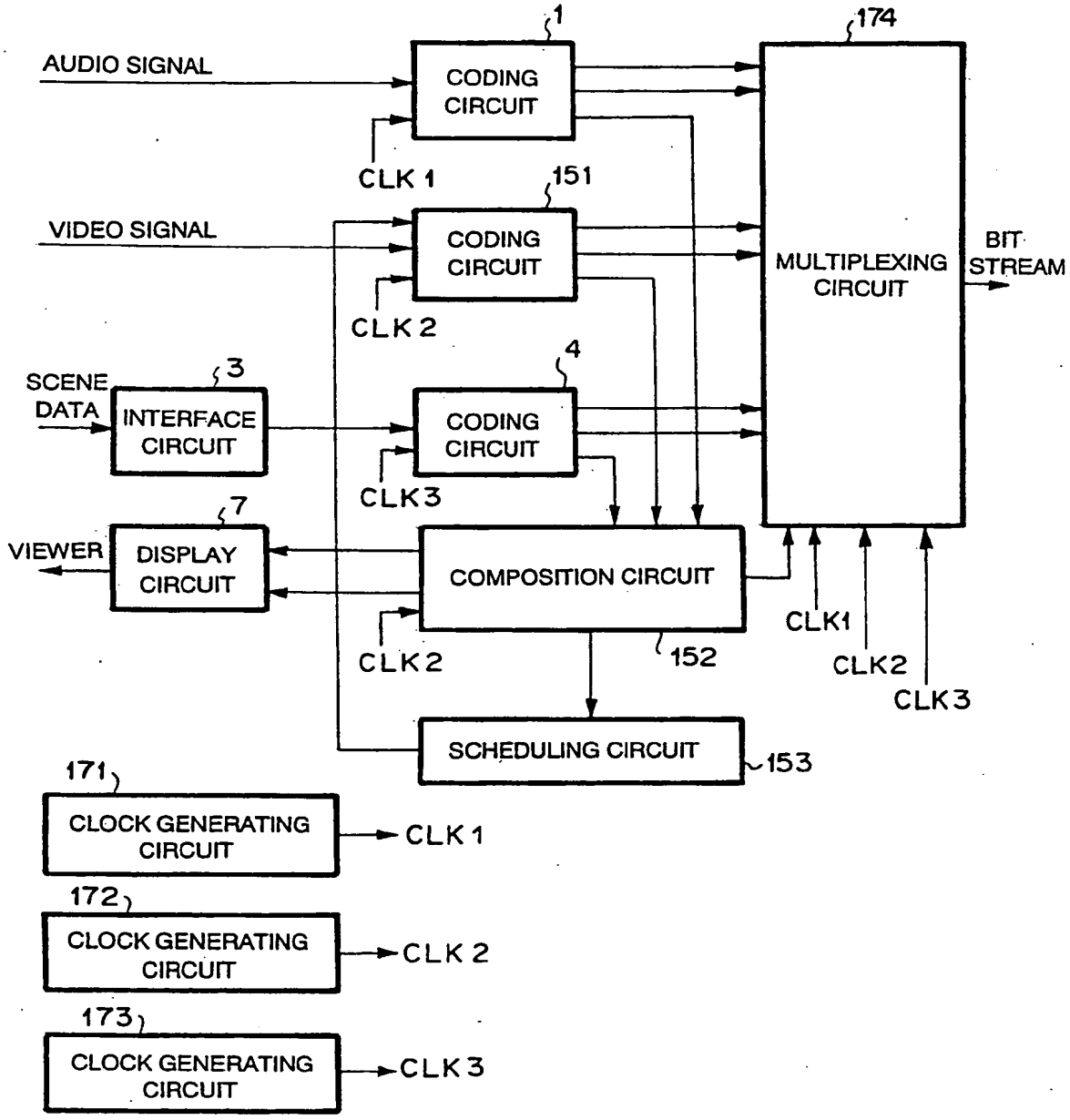
BNSDOCID: <EP_____0924934A1_I_>

FIG.15



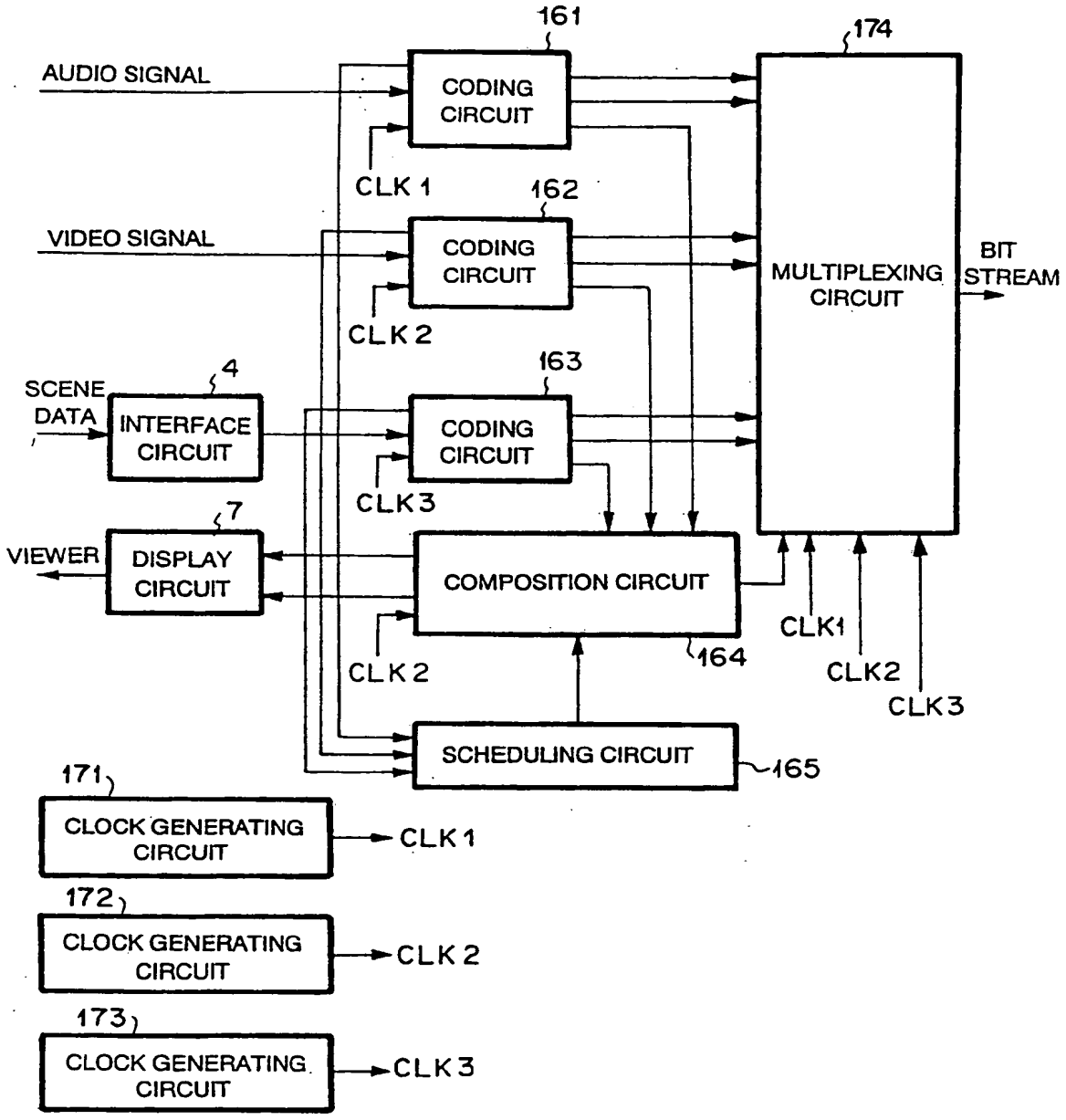
BNSDOCID: <EP_____0924934A1_I_>

FIG.16



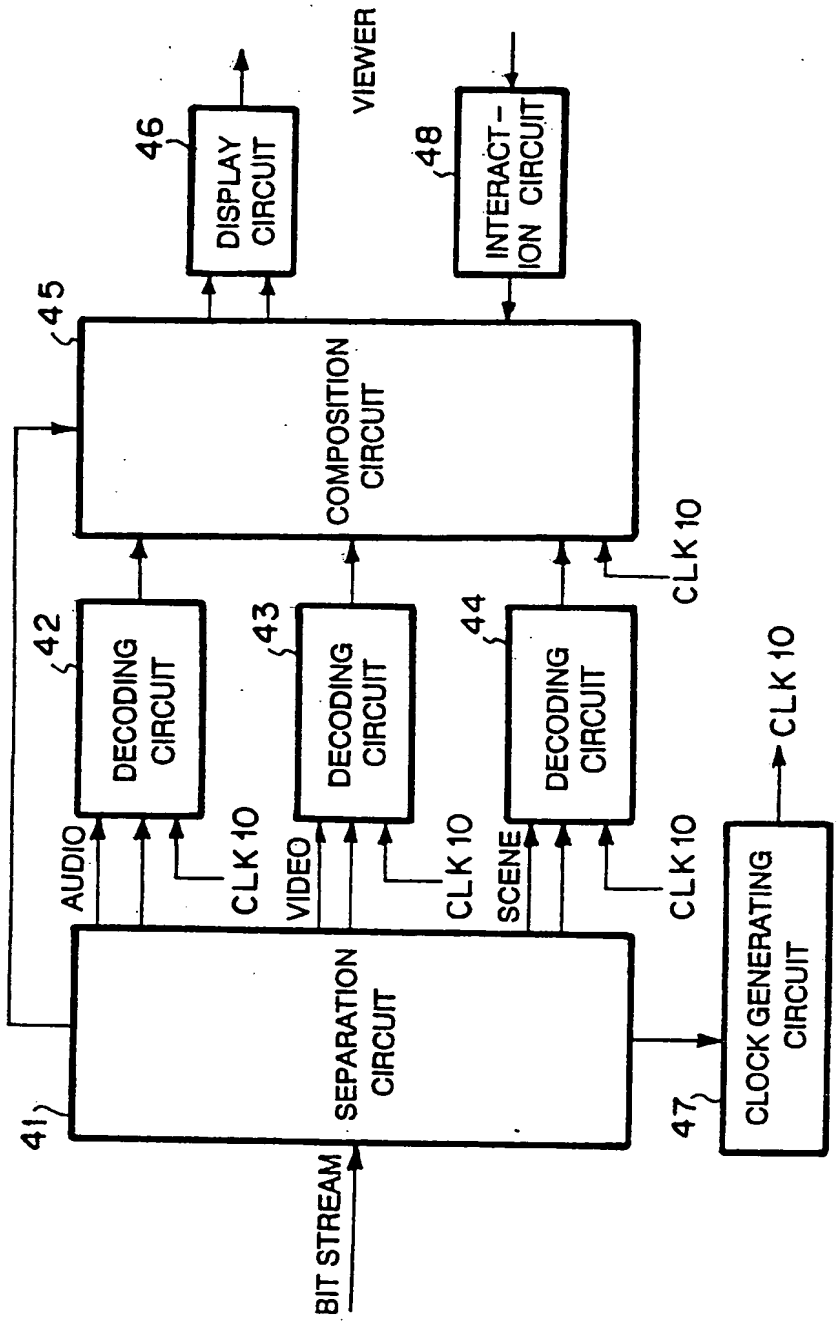
BNSDOCID: <EP_____0924934A1_I_>

FIG.17



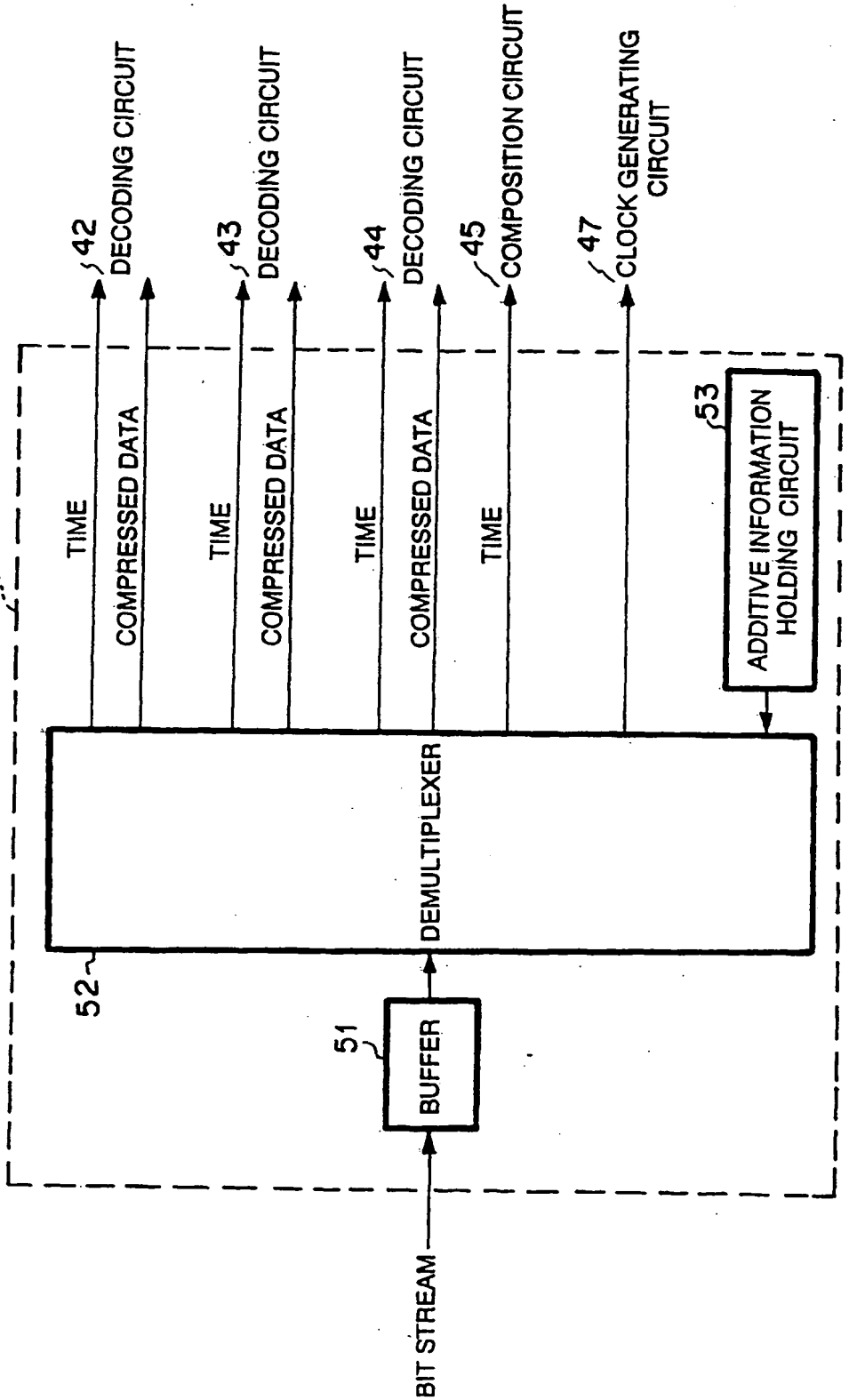
3NSDOCID: <EP_____0924934A1_I_>

FIG.18



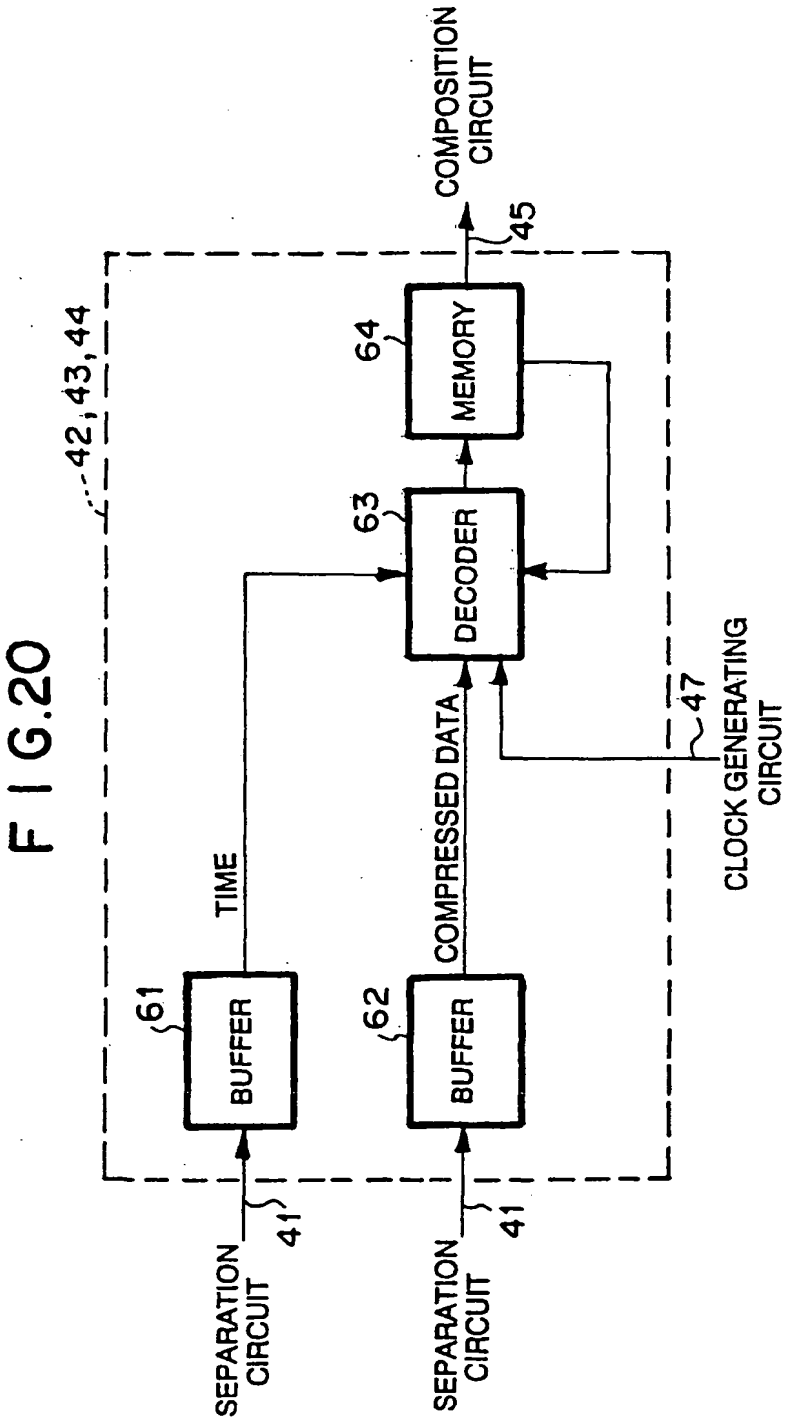
BNSDOCID: <EP_____0924934A1_I_>

FIG.19



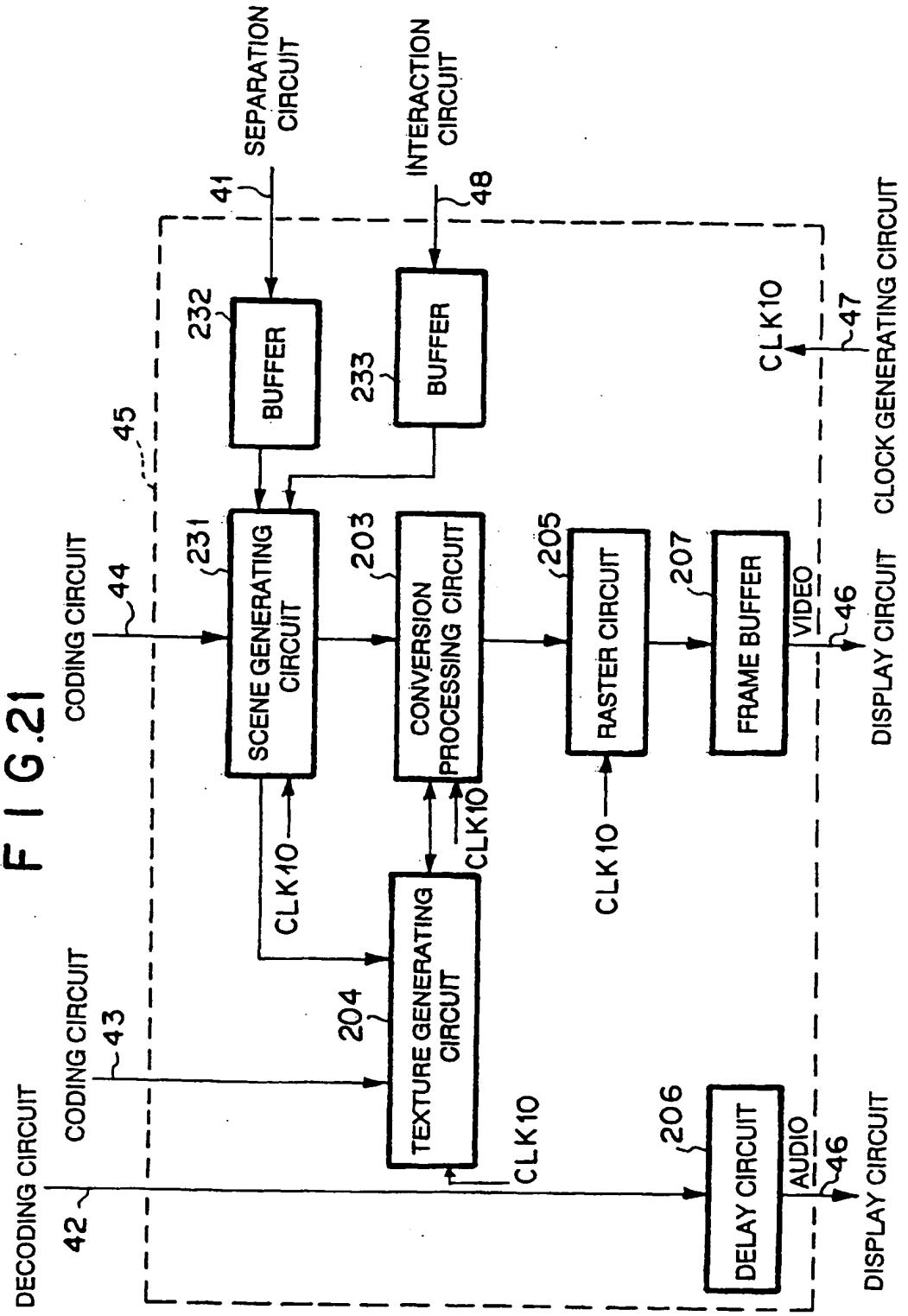
BNSDOCID: <EP_____0924934A1_I_>

FIG.20



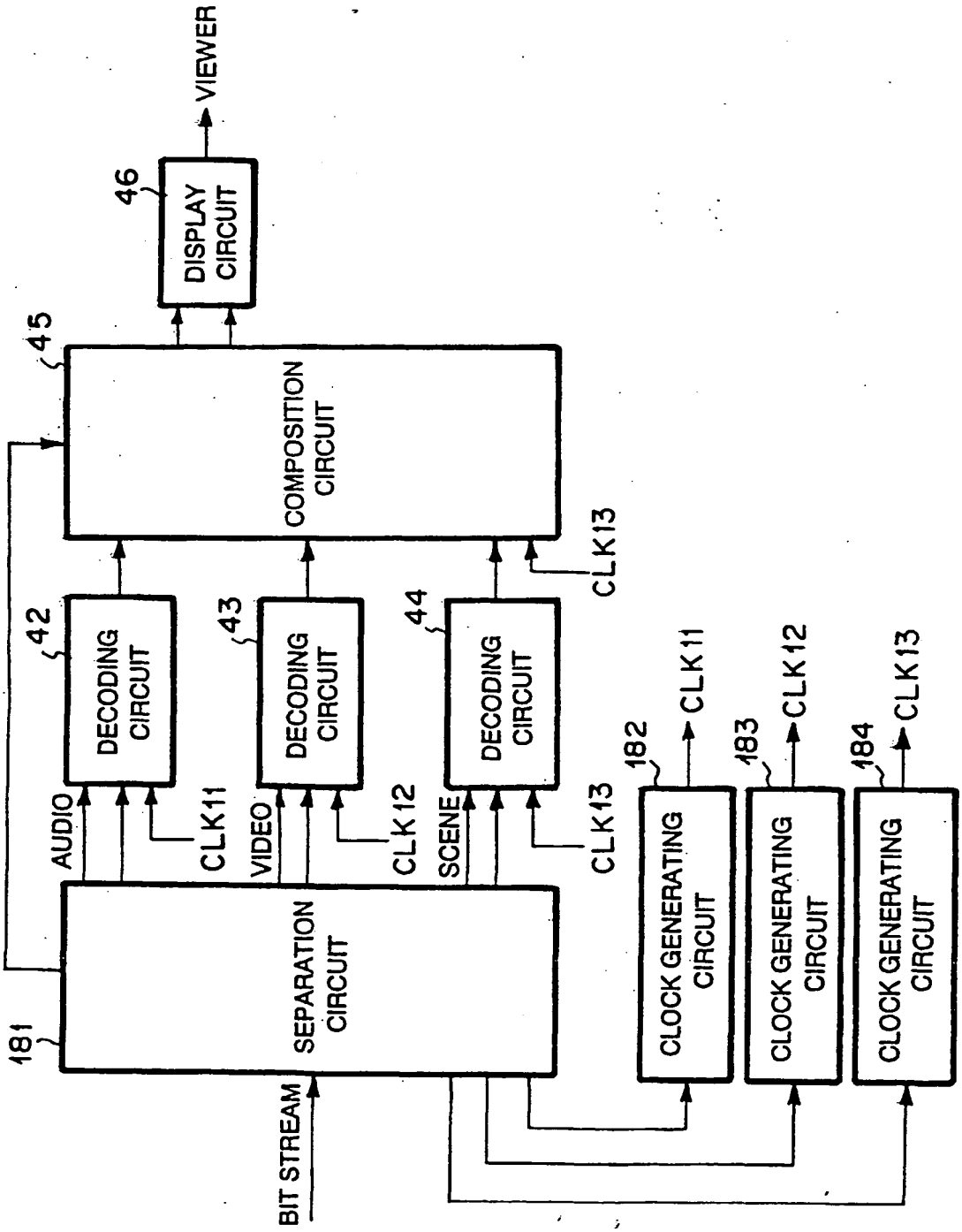
BNSDOCID: <EP_____0924934A1_I_>

FIG.21



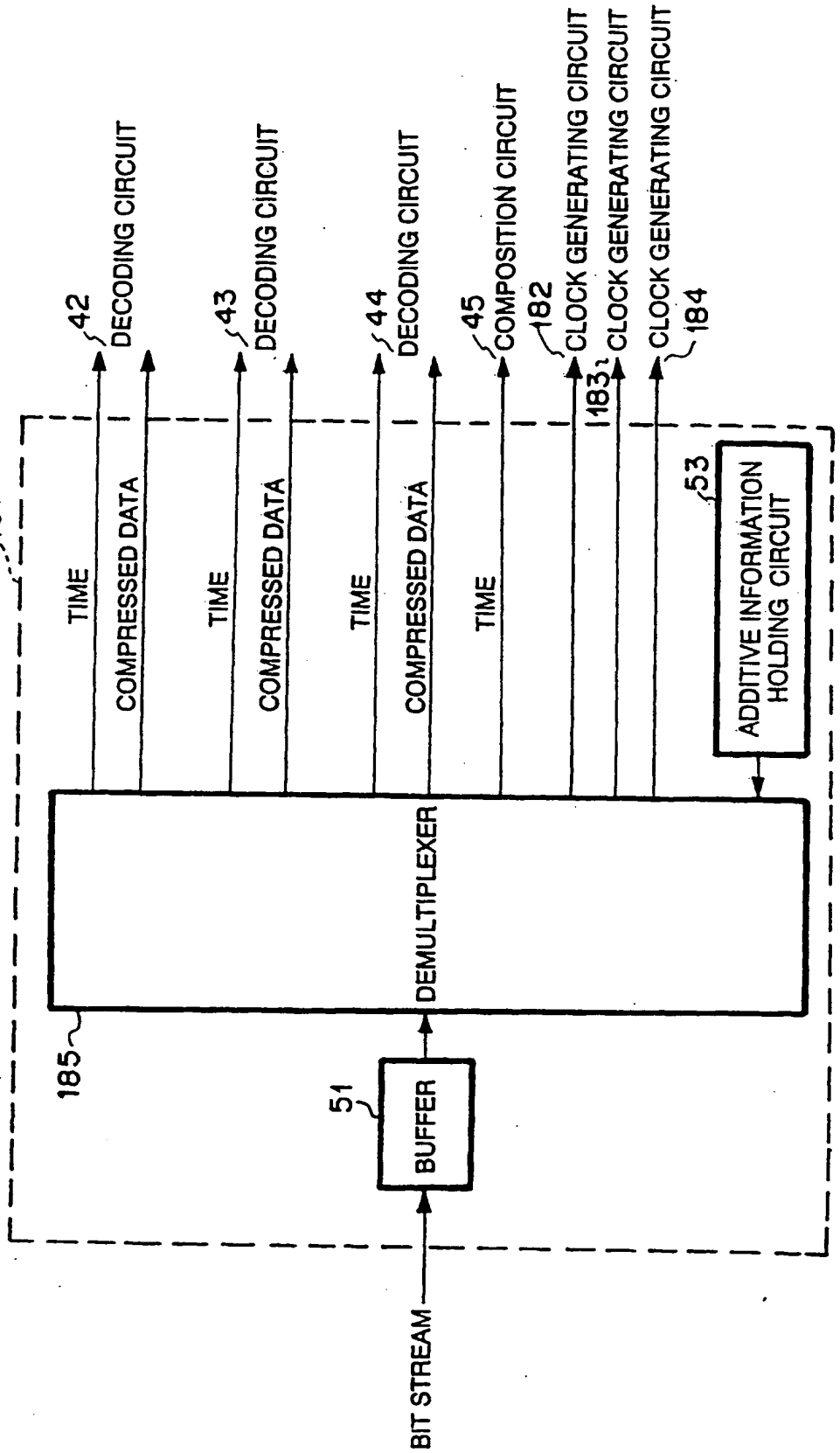
BNSDOCID: <EP_____0924934A1_I_>

FIG.22



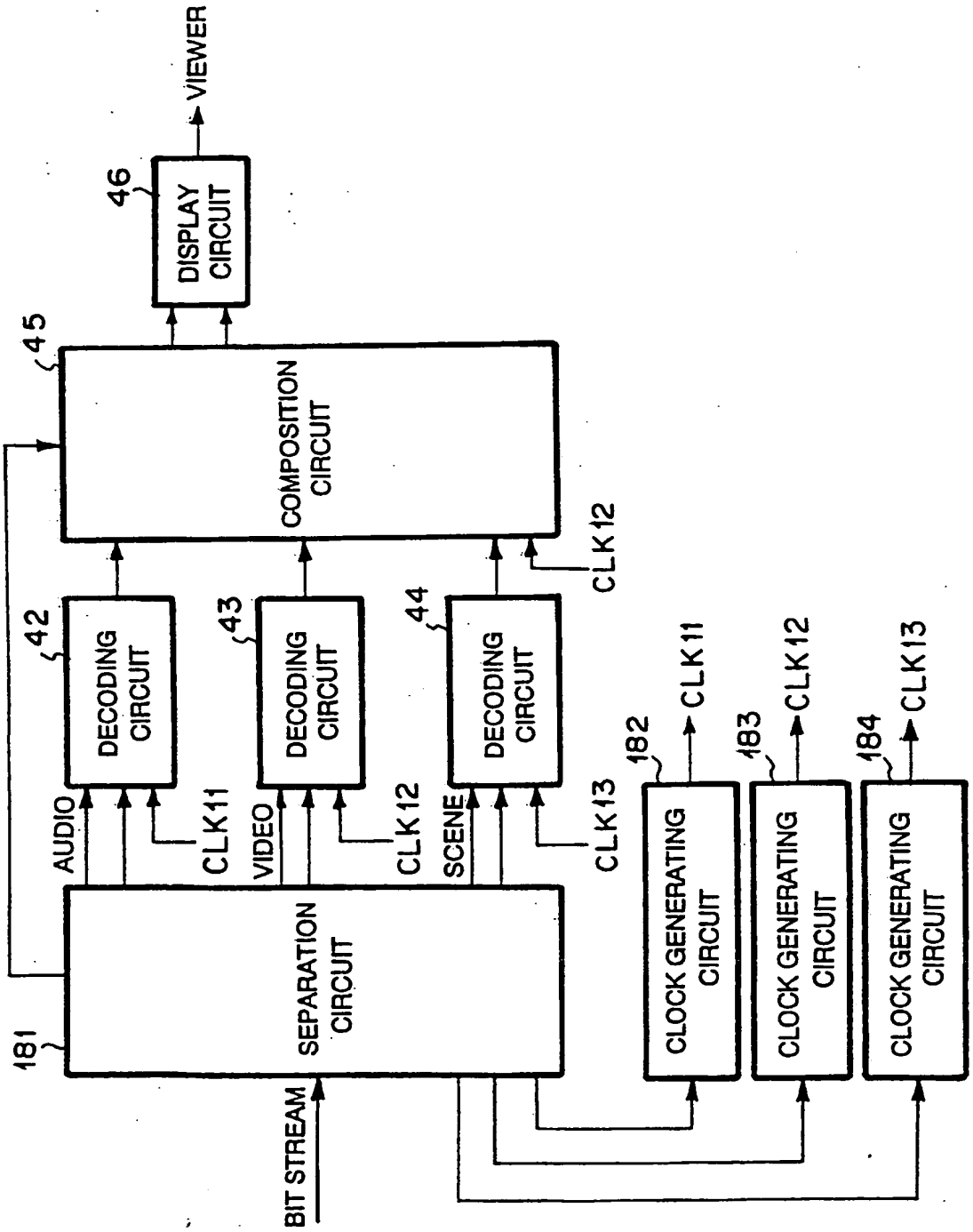
BNSDOCID: <EP_____0924934A1_I_>

FIG.23



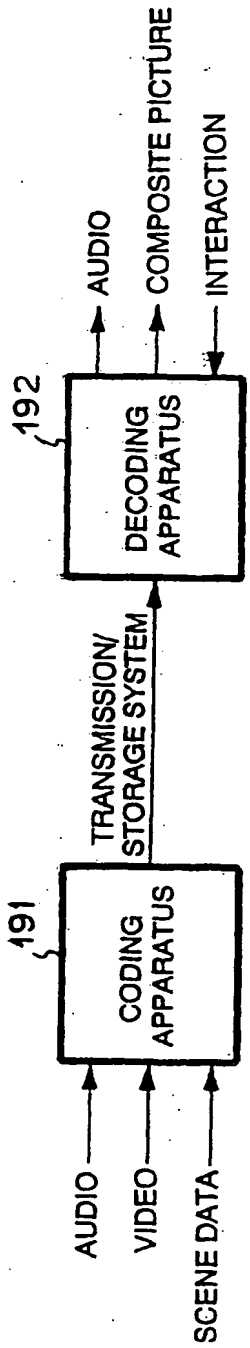
BNSDOCID: <EP_____0924934A1_I_>

FIG.24



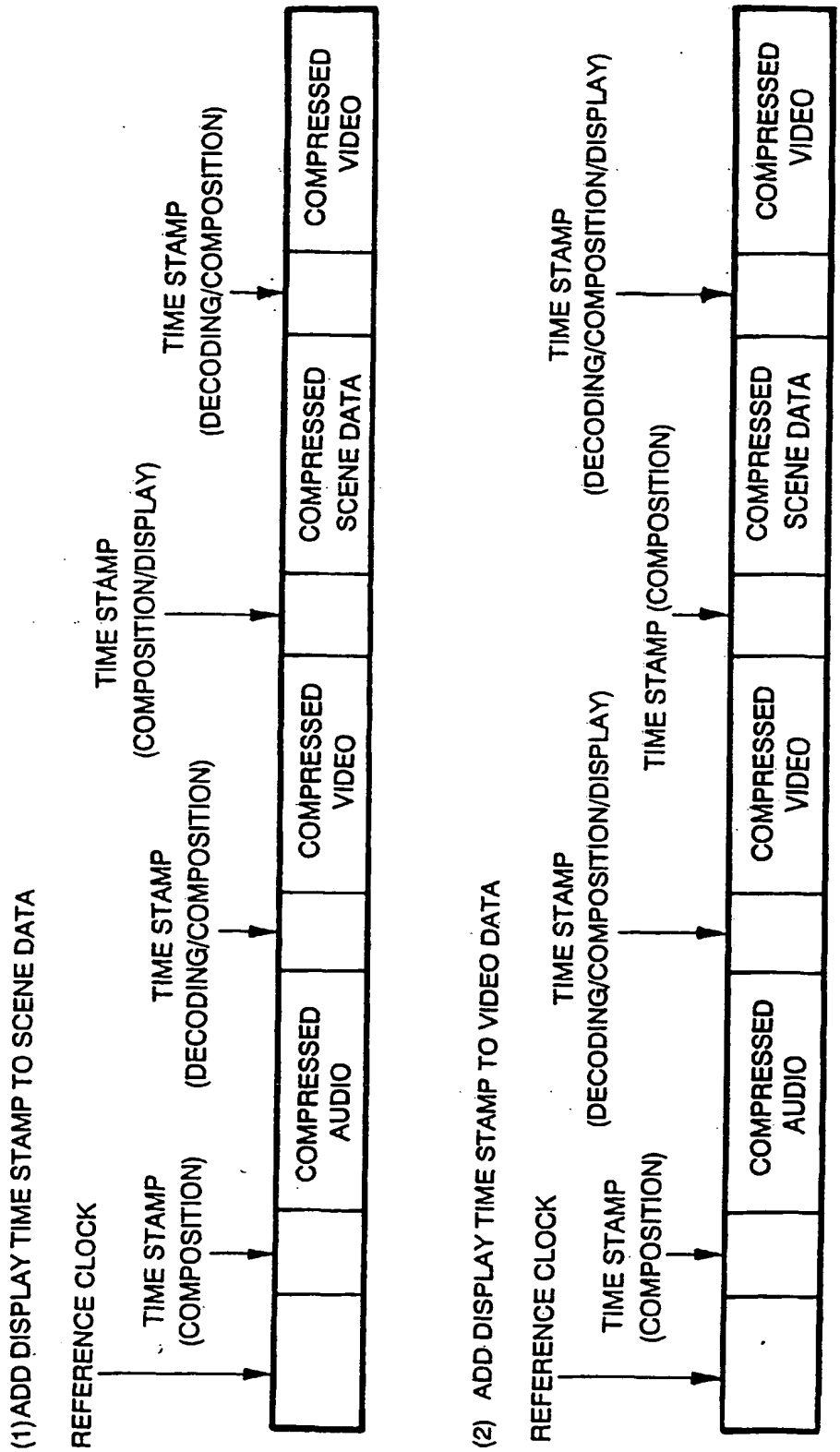
BNSDOCID: <EP_____0924934A1_I_>

FIG.25



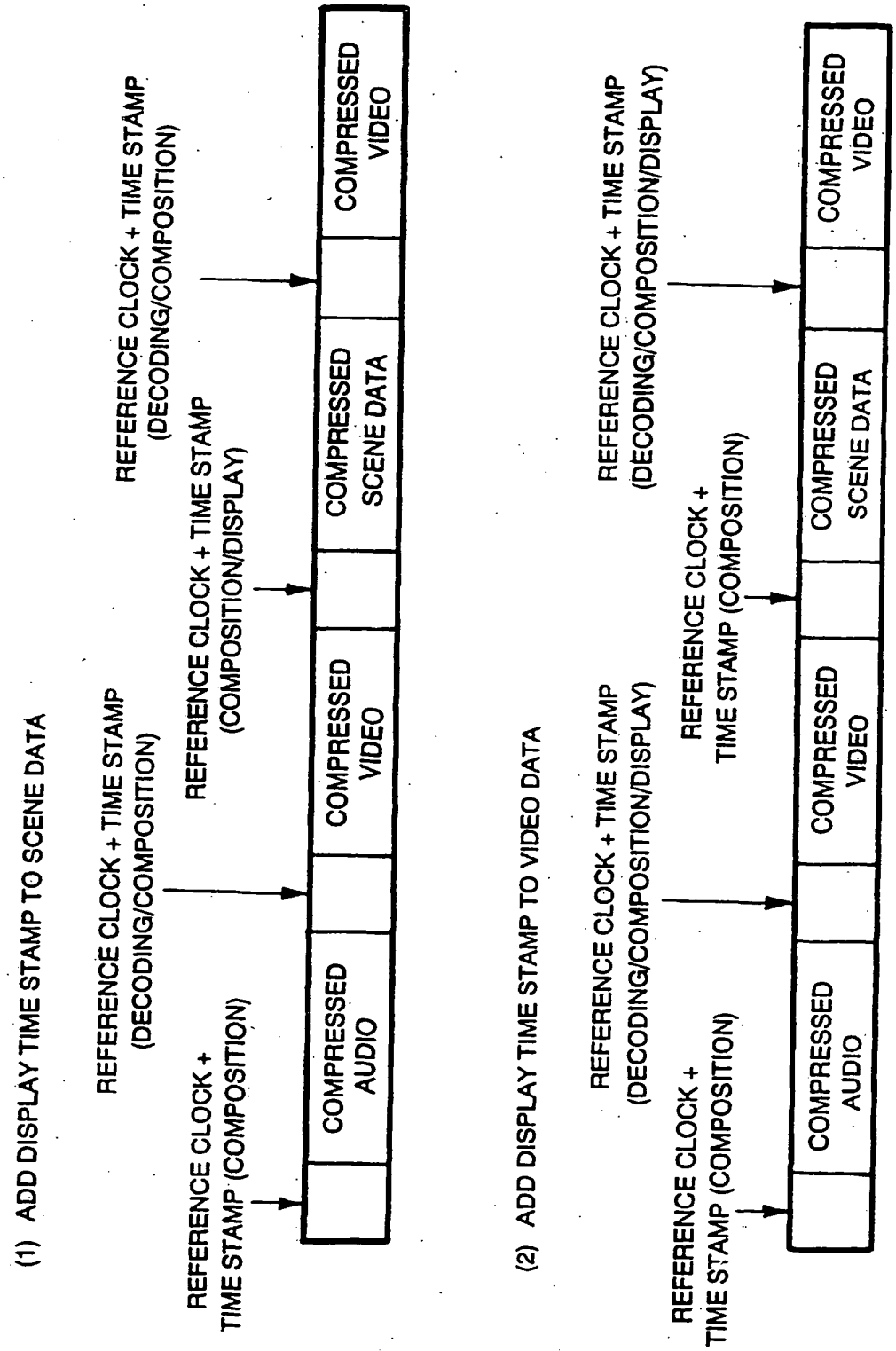
BNSDOCID: <EP_____0924934A1_I_>

FIG.26



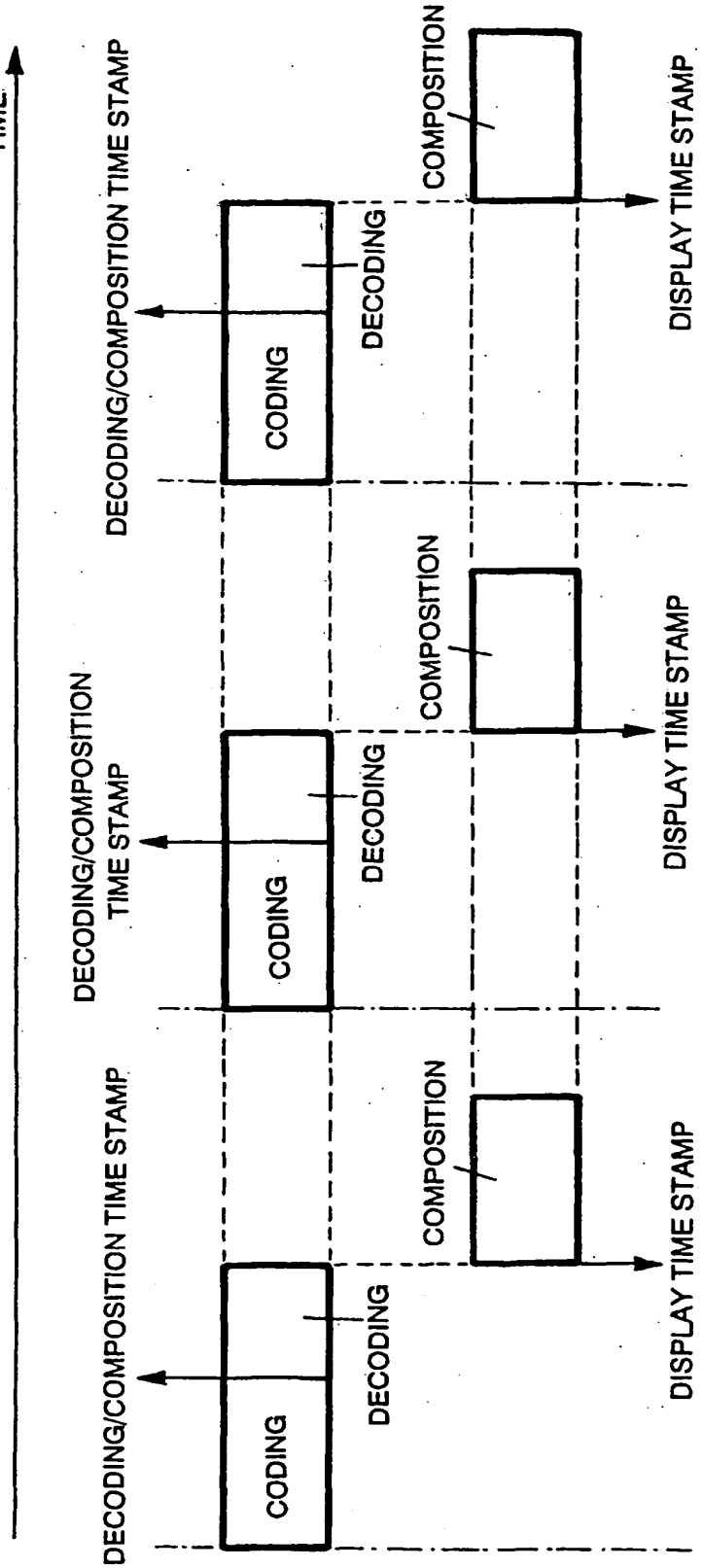
BNSDOCID: <EP_____0924934A1_I_>

FIG.27



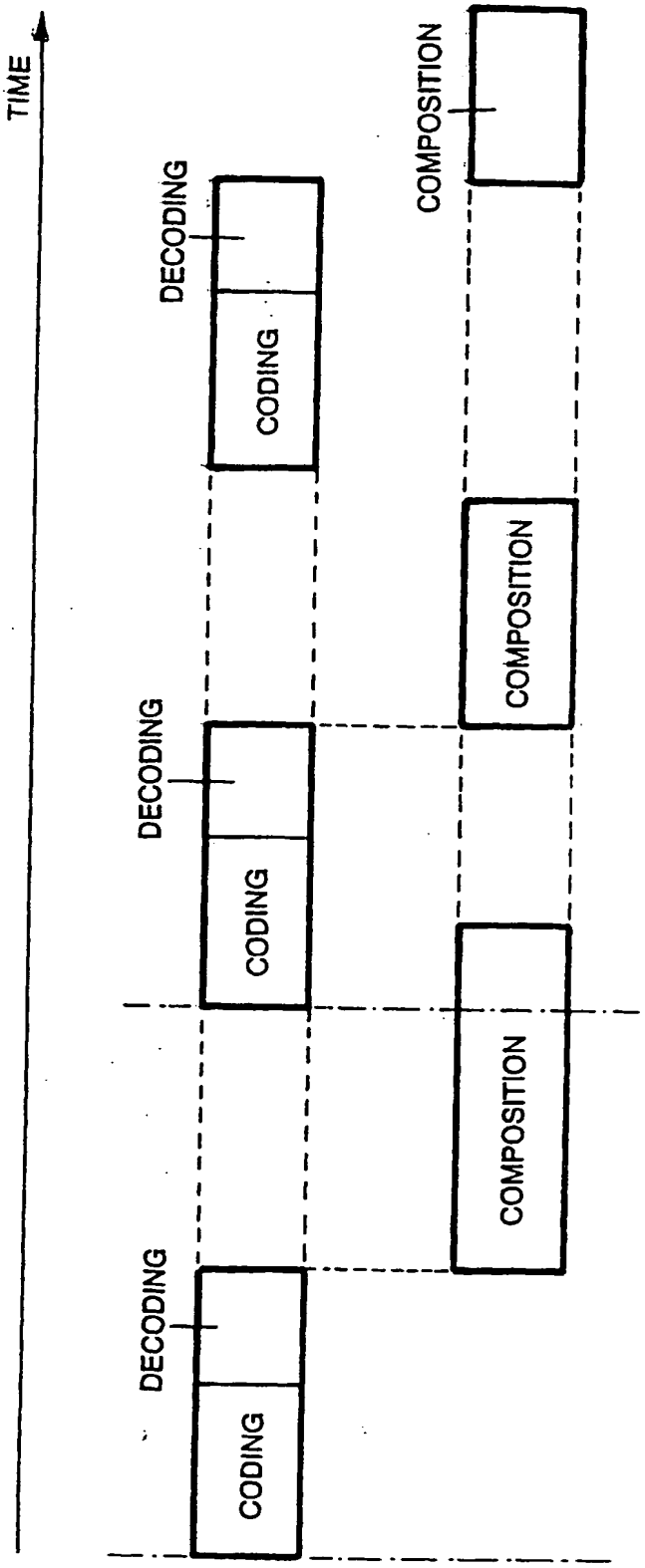
3NSDOCID: <EP_____0924934A1_I_>

FIG.28



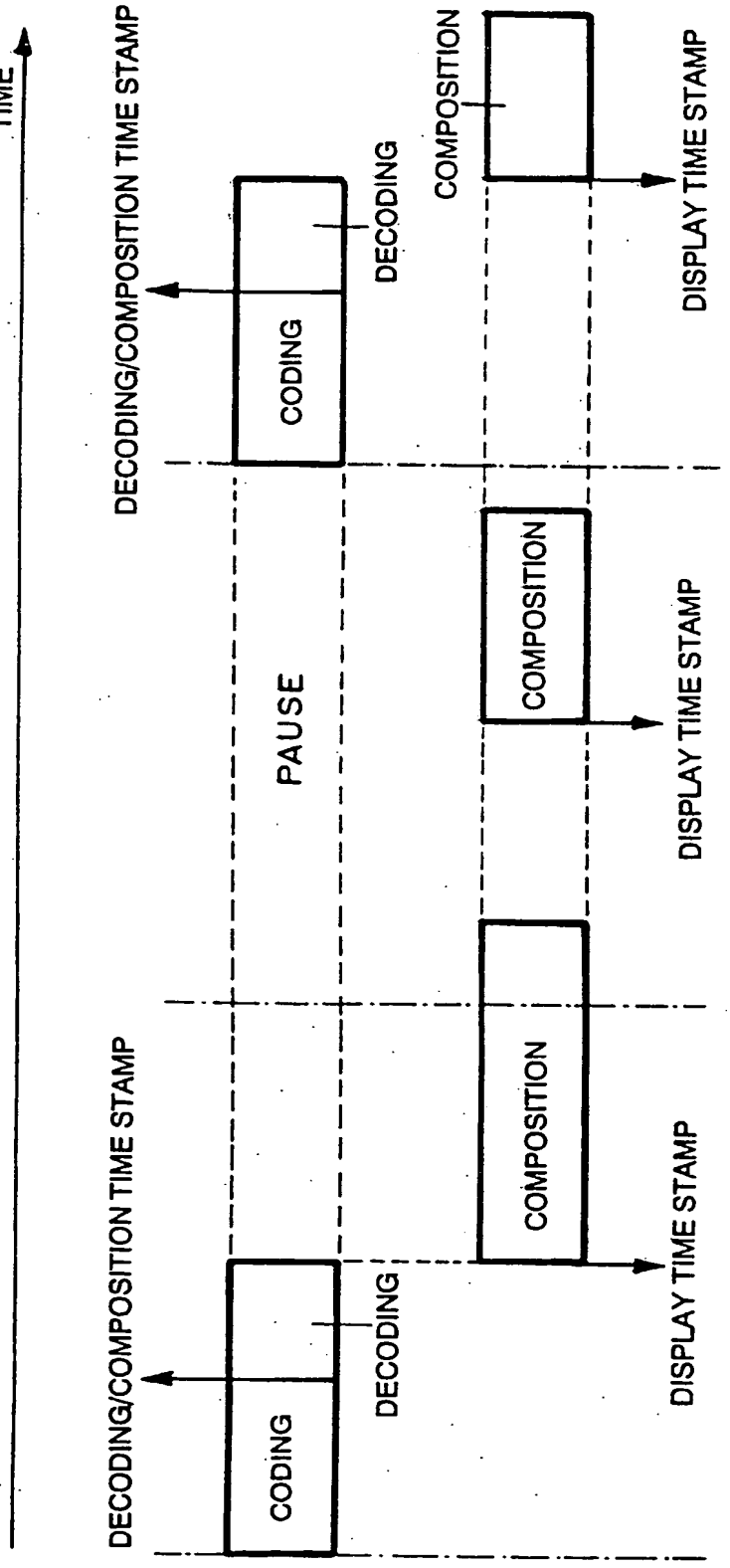
BNSDOCID: <EP_____0924934A1_I_>

FIG.29



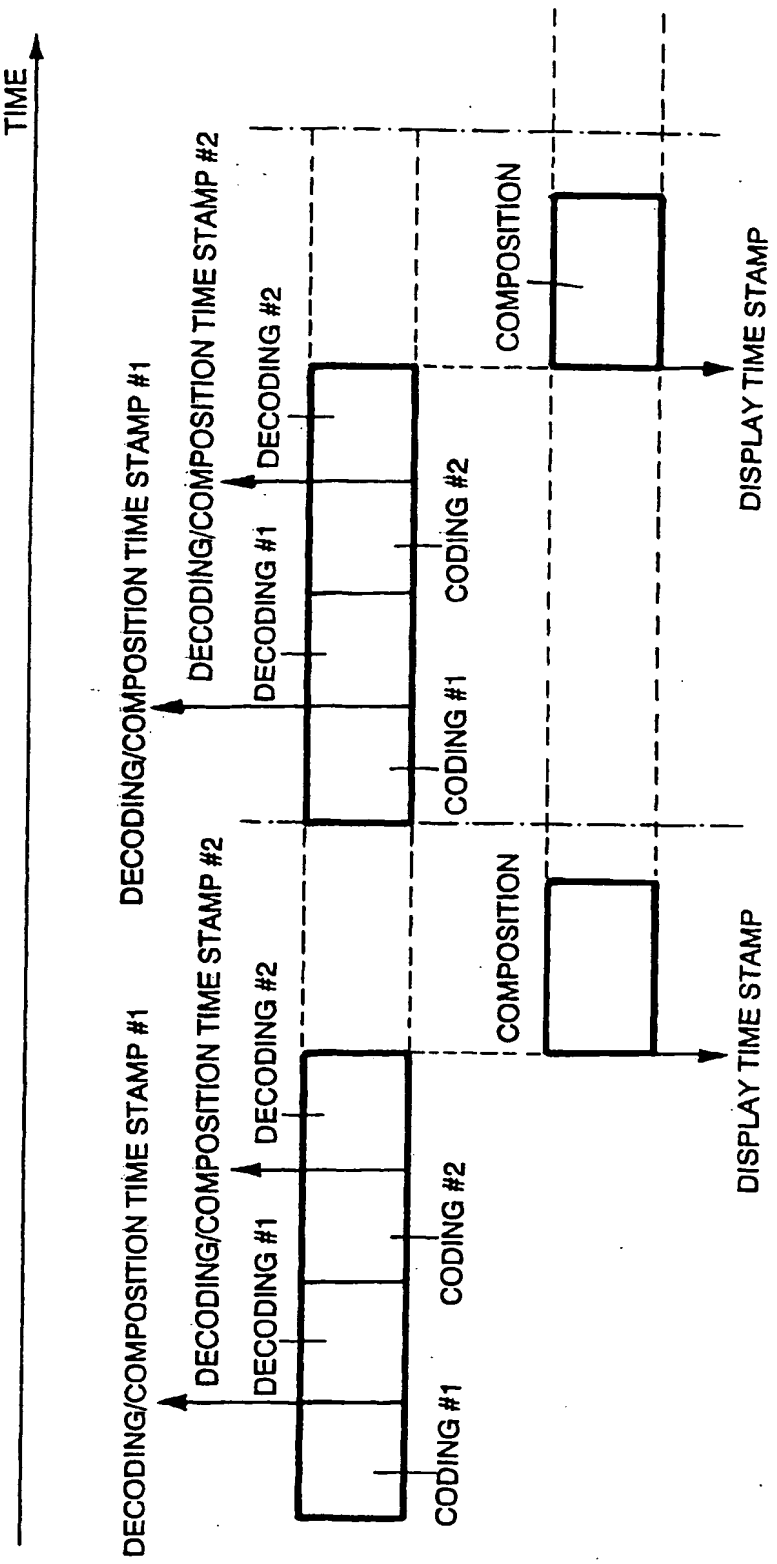
BNSDOCID: <EP_____0924934A1_I_>

FIG.30



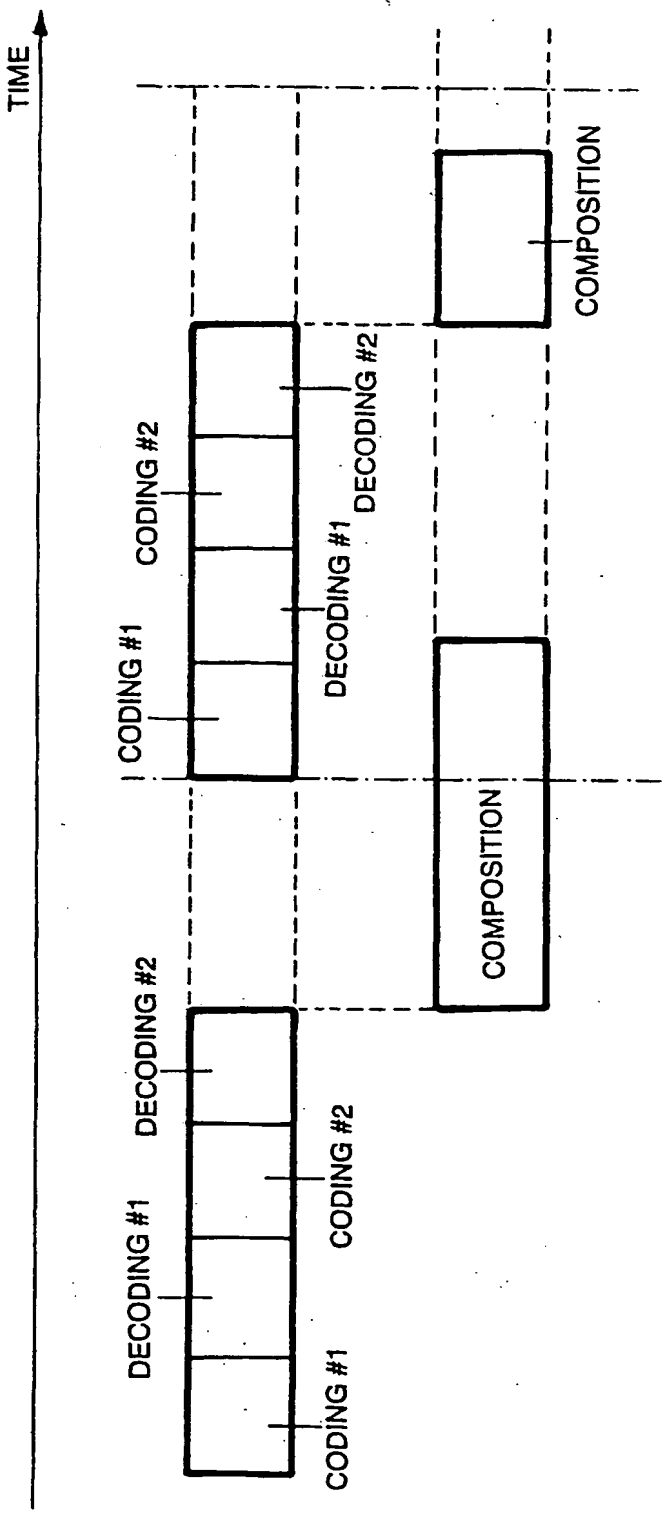
BNSDOCID: <EP_____0924934A1_I_>

FIG.31



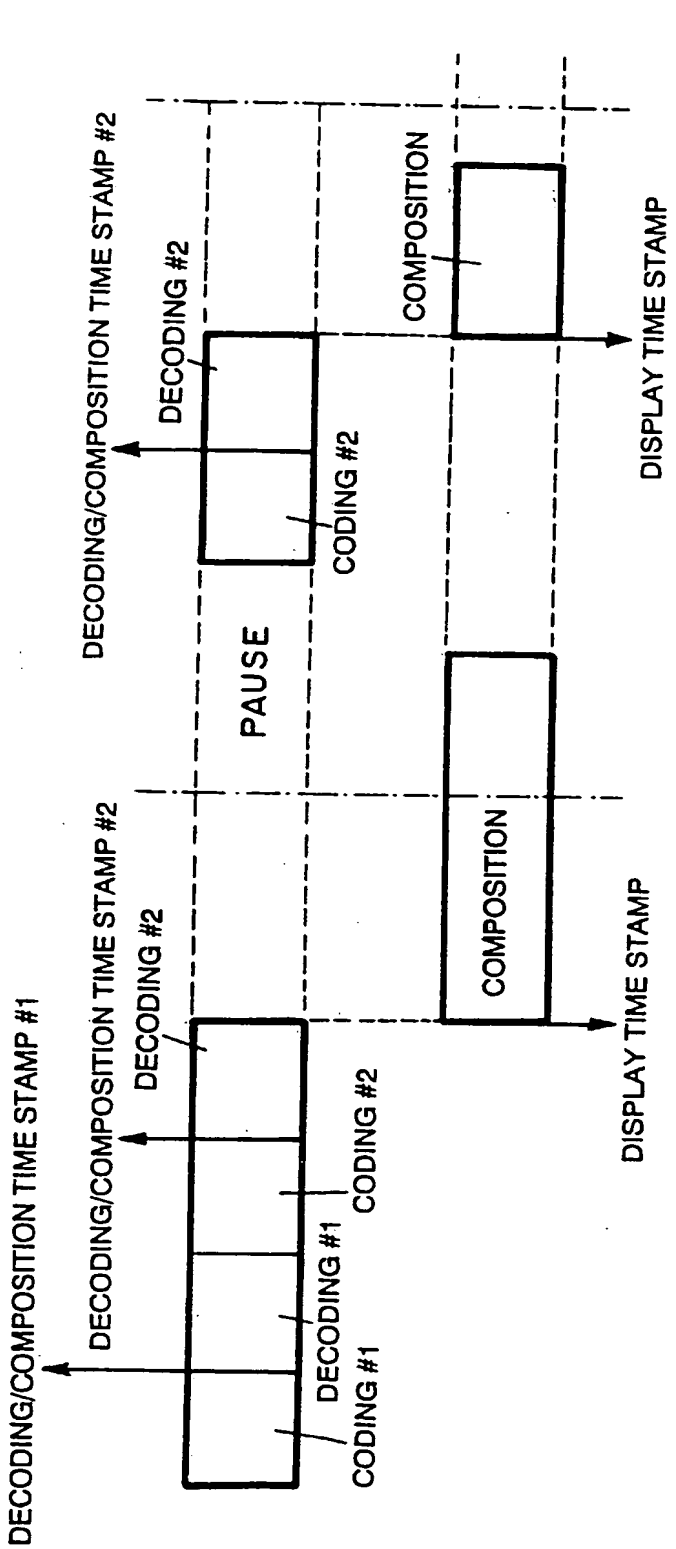
BNSDOCID: <EP_____0924934A1_I_>

FIG.32



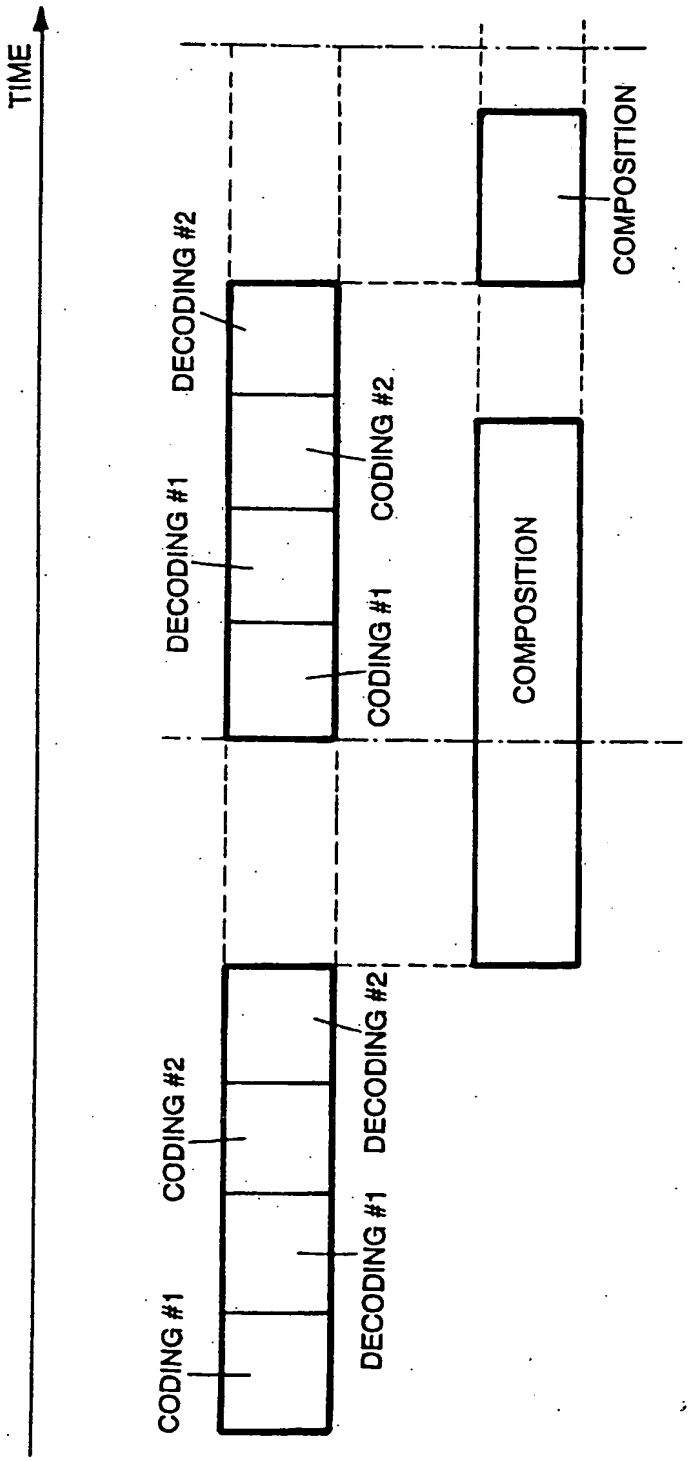
BNSDOCID: <EP_____0924934A1_I_>

FIG.33



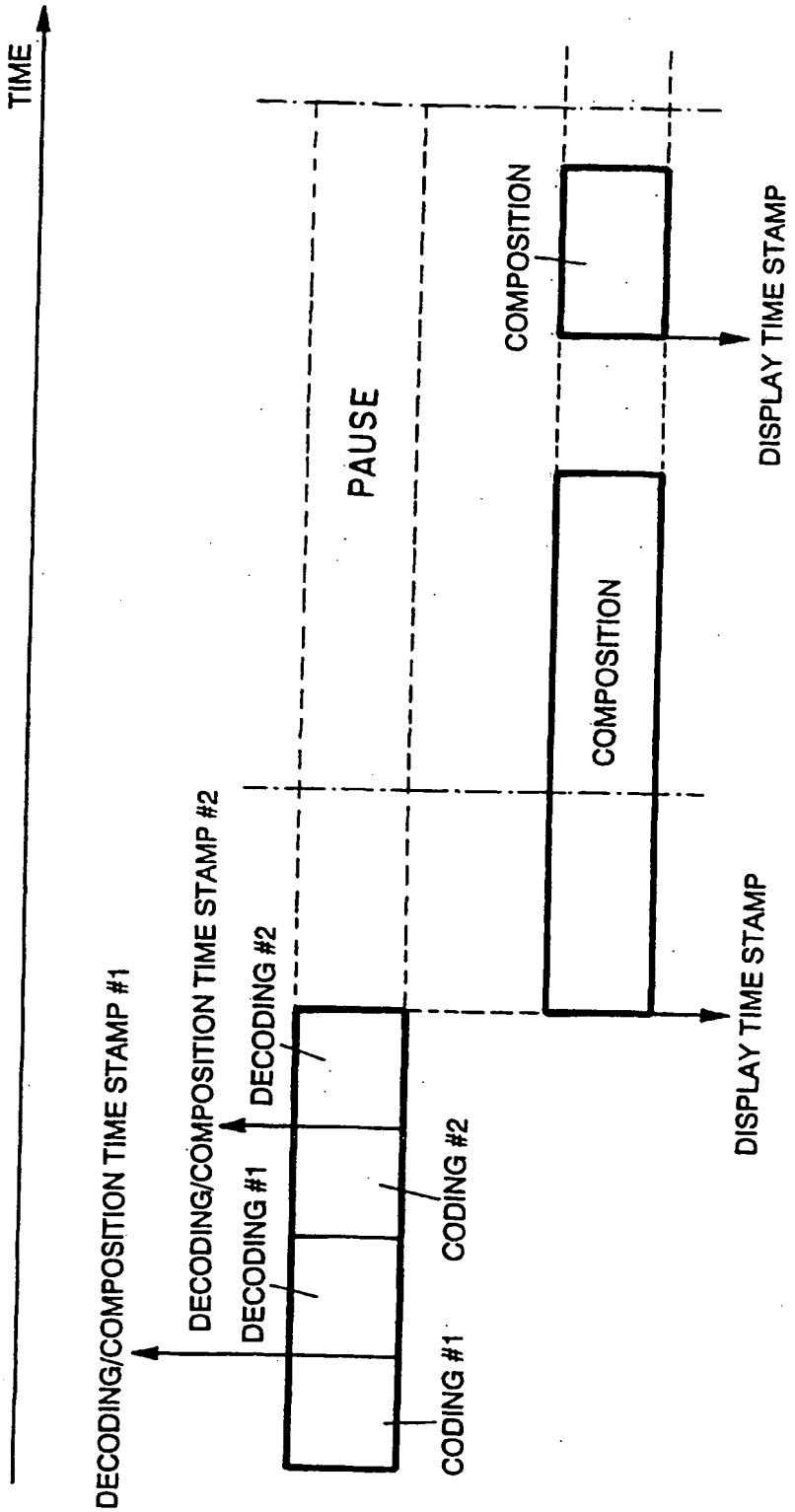
BNSDOCID: <EP_____0924934A1_I_>

FIG.34



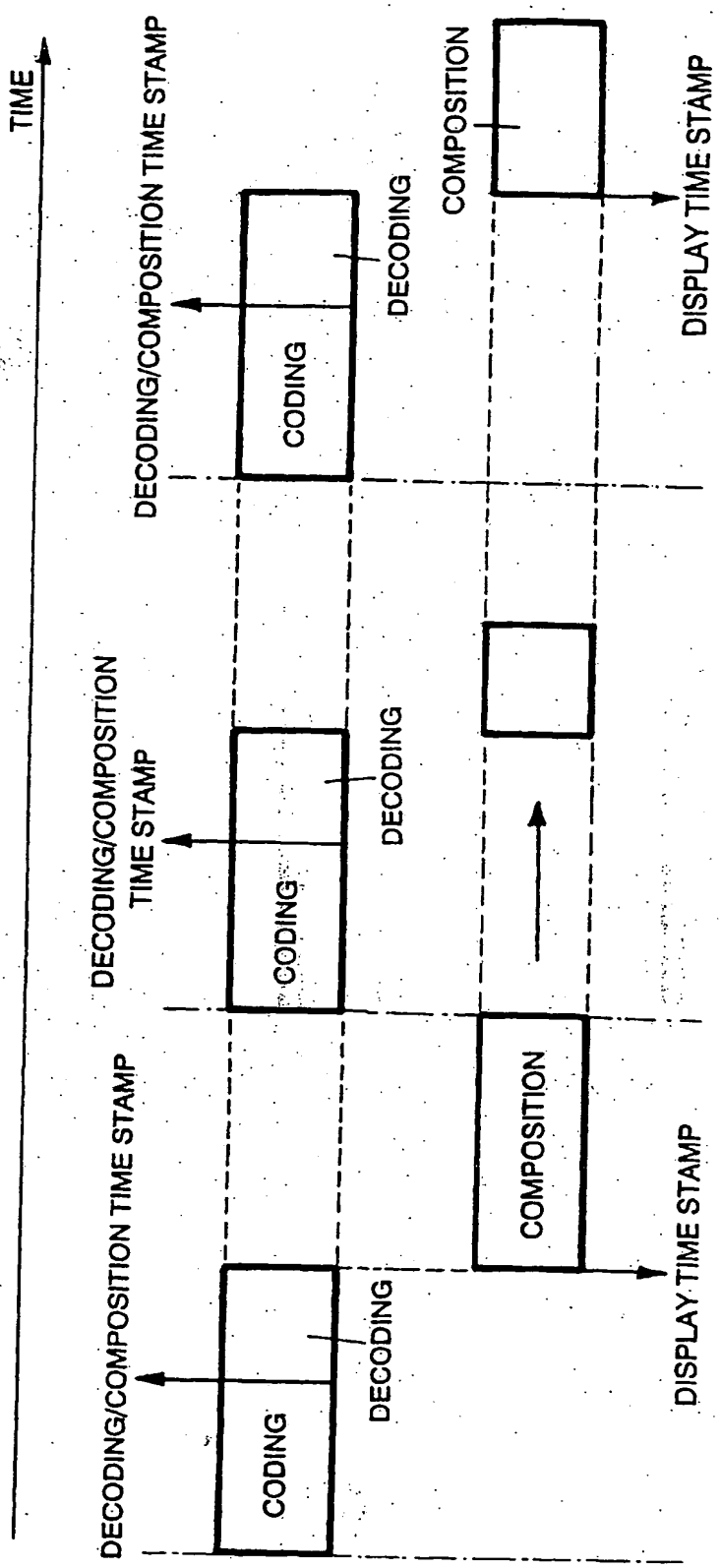
BNSDOCID: <EP_____0924934A1_I_>

FIG.35



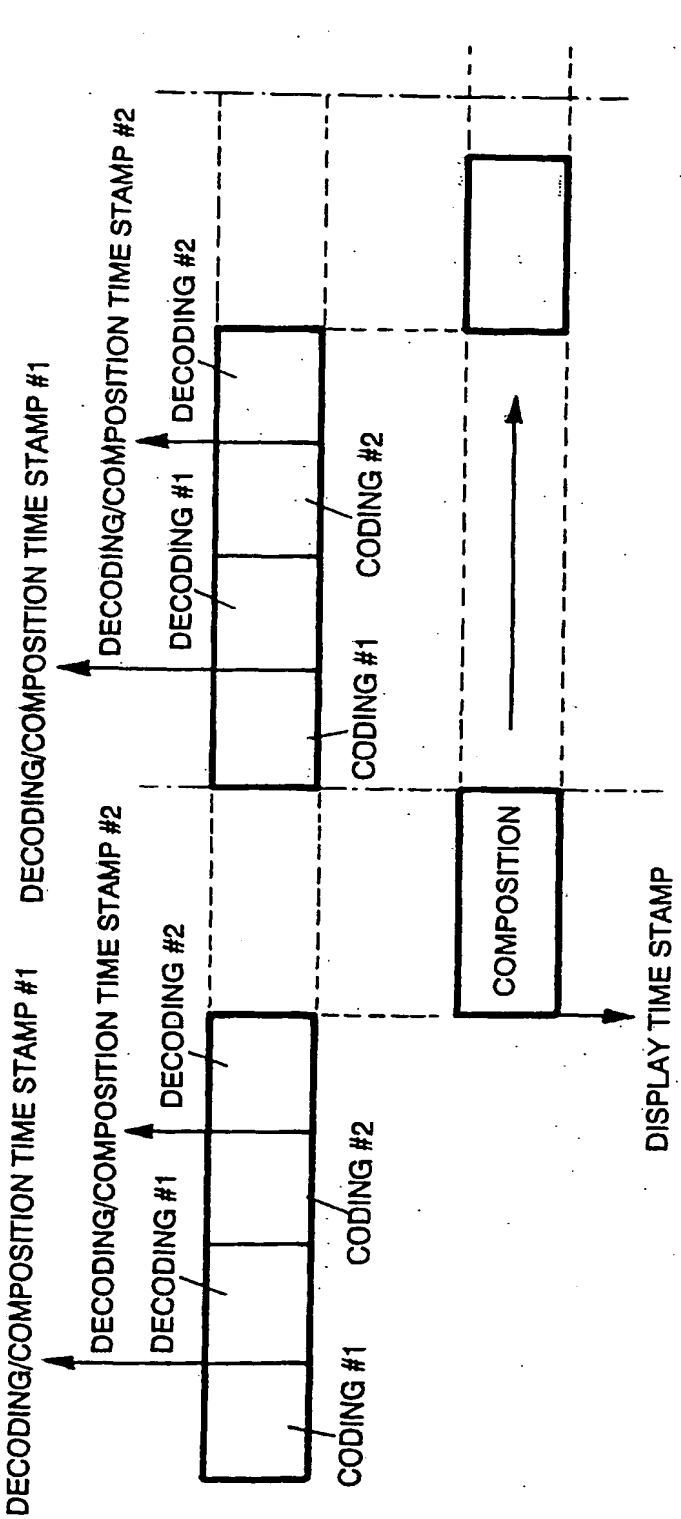
BNSDOCID: <EP_____0924934A1_I_>

FIG.36



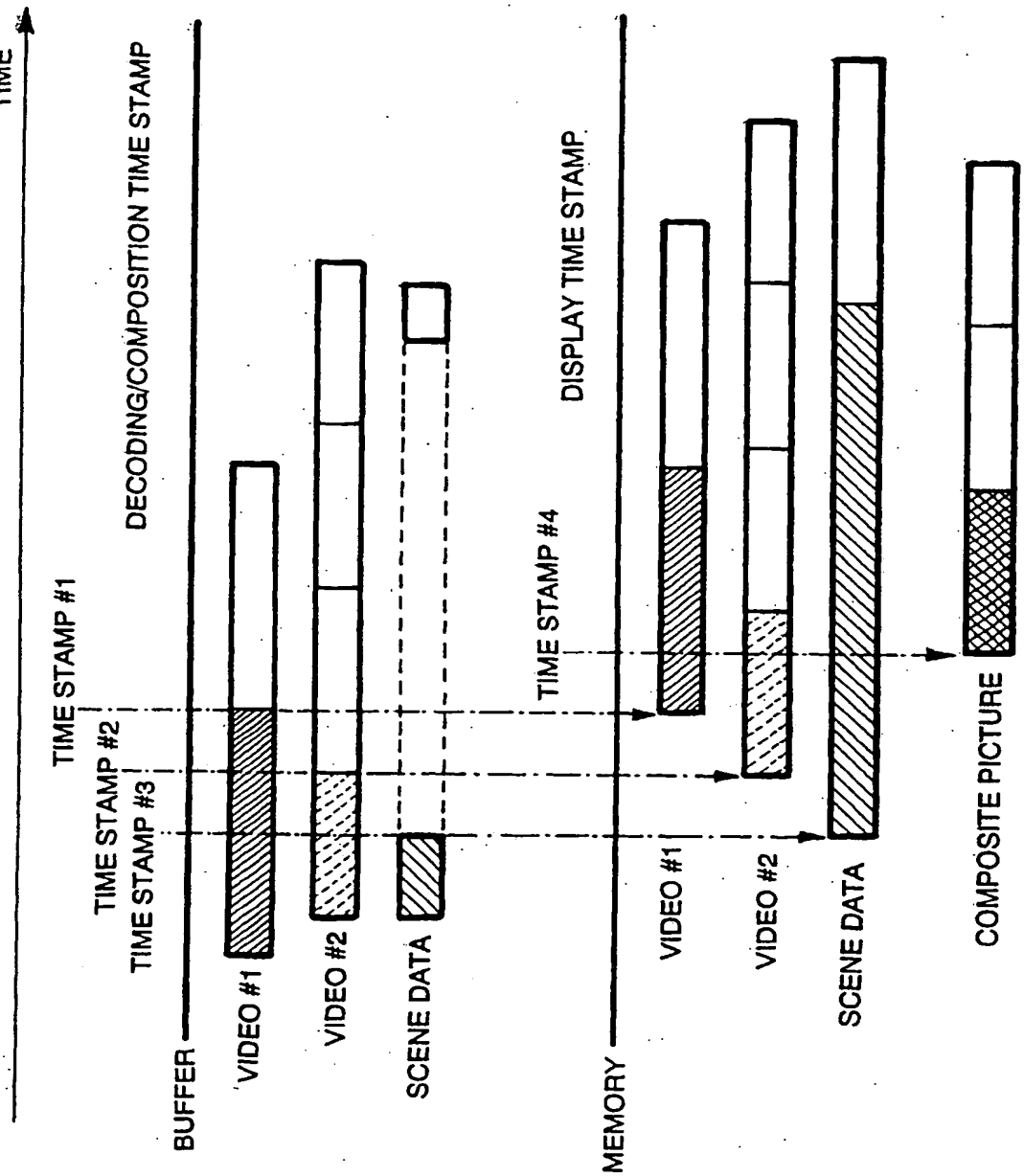
BNSDOCID: <EP_____0924934A1_i_>

FIG.37



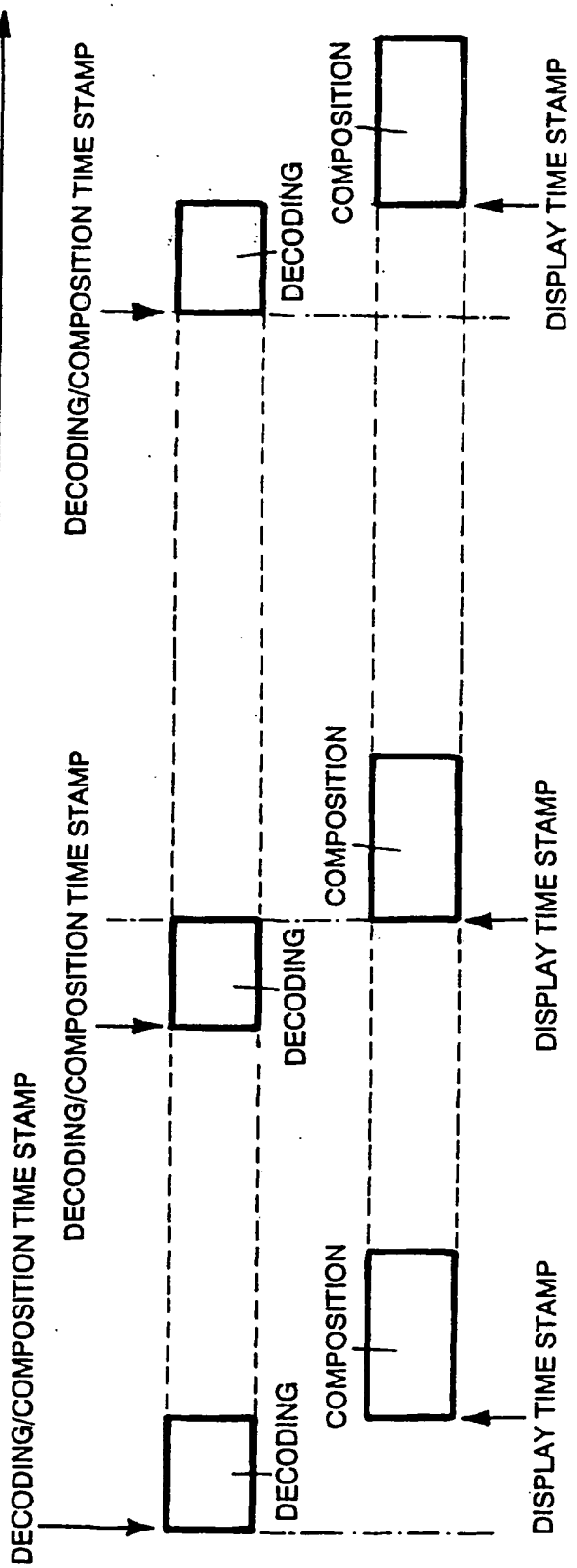
BNSDOCID: <EP_____0924934A1_I_>

FIG.38



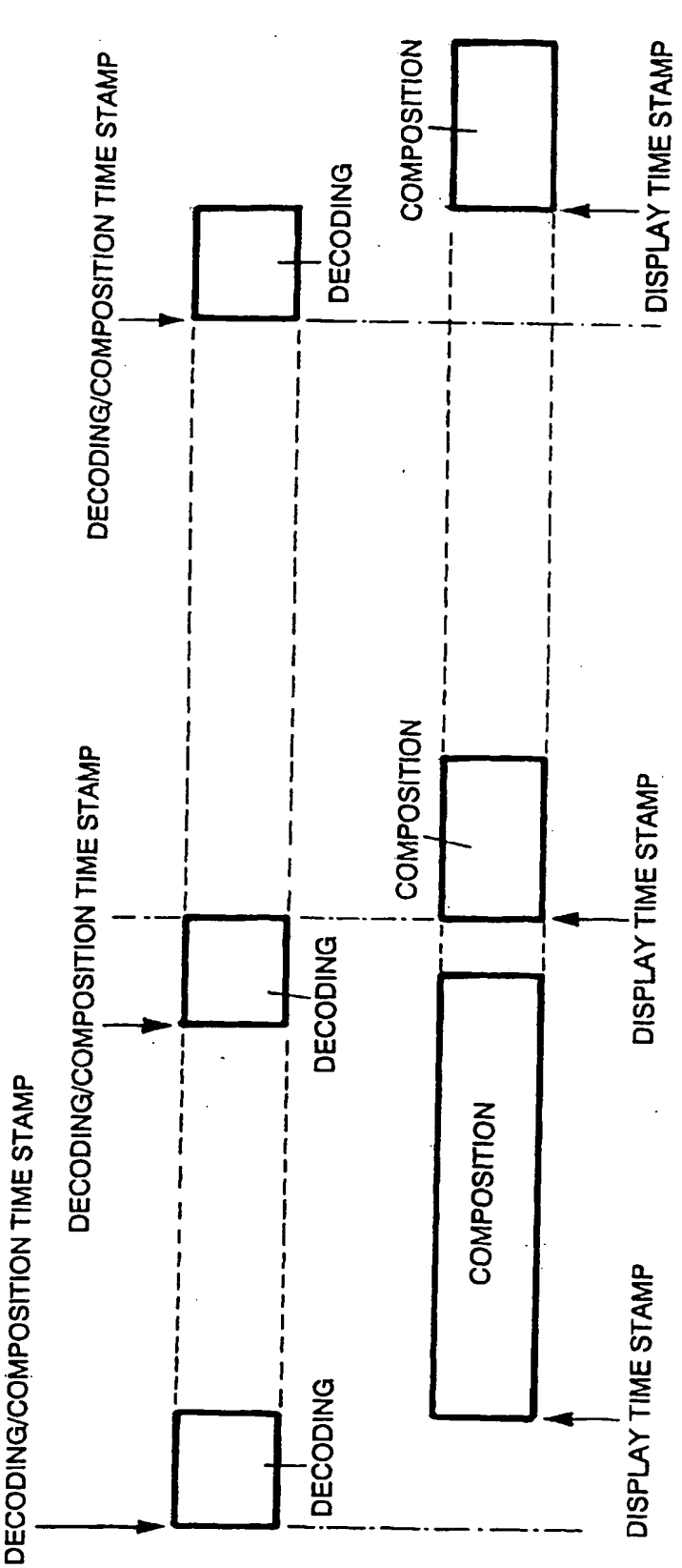
BNSDOCID: <EP_____0924934A1_I_>

FIG.39



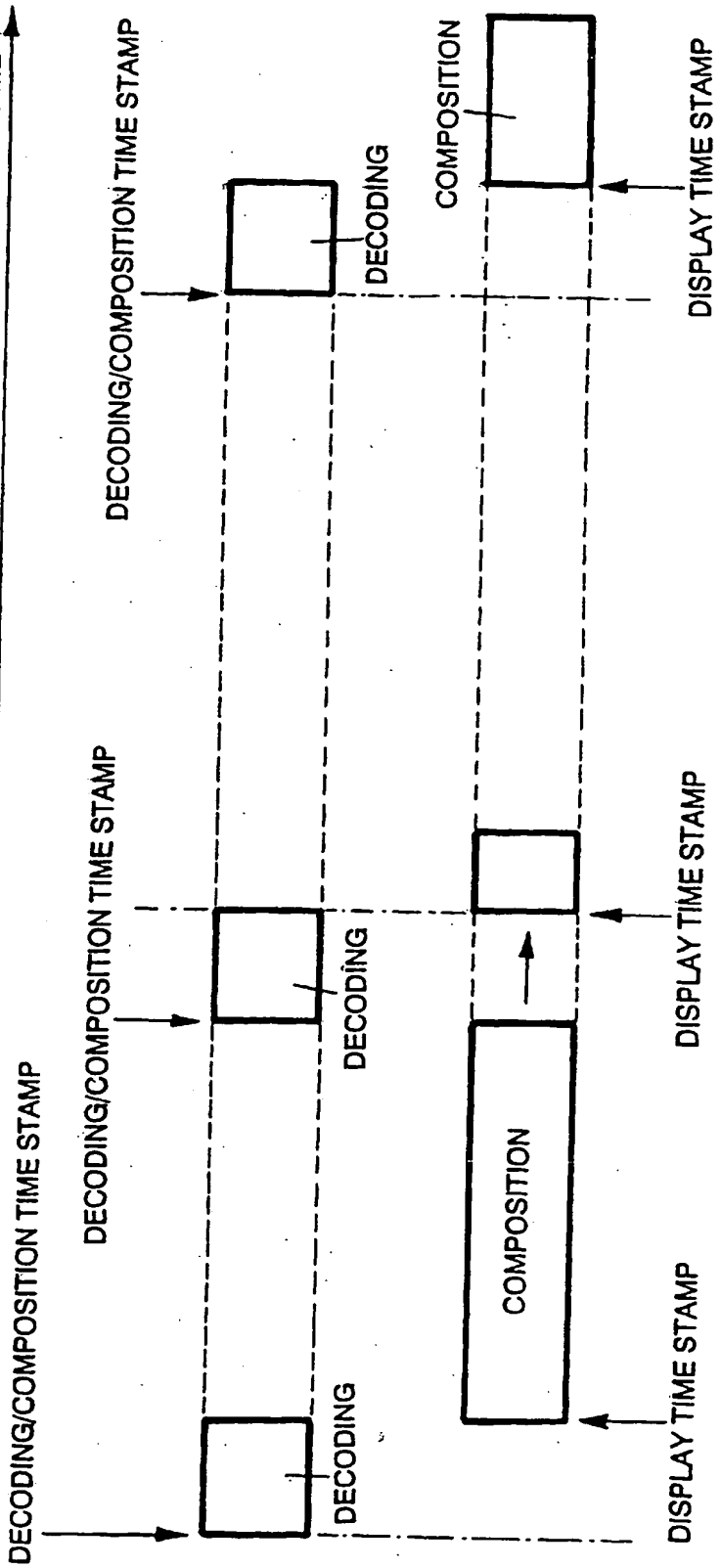
BNSDOCID: <EP_____0924934A1_I_>

FIG.40



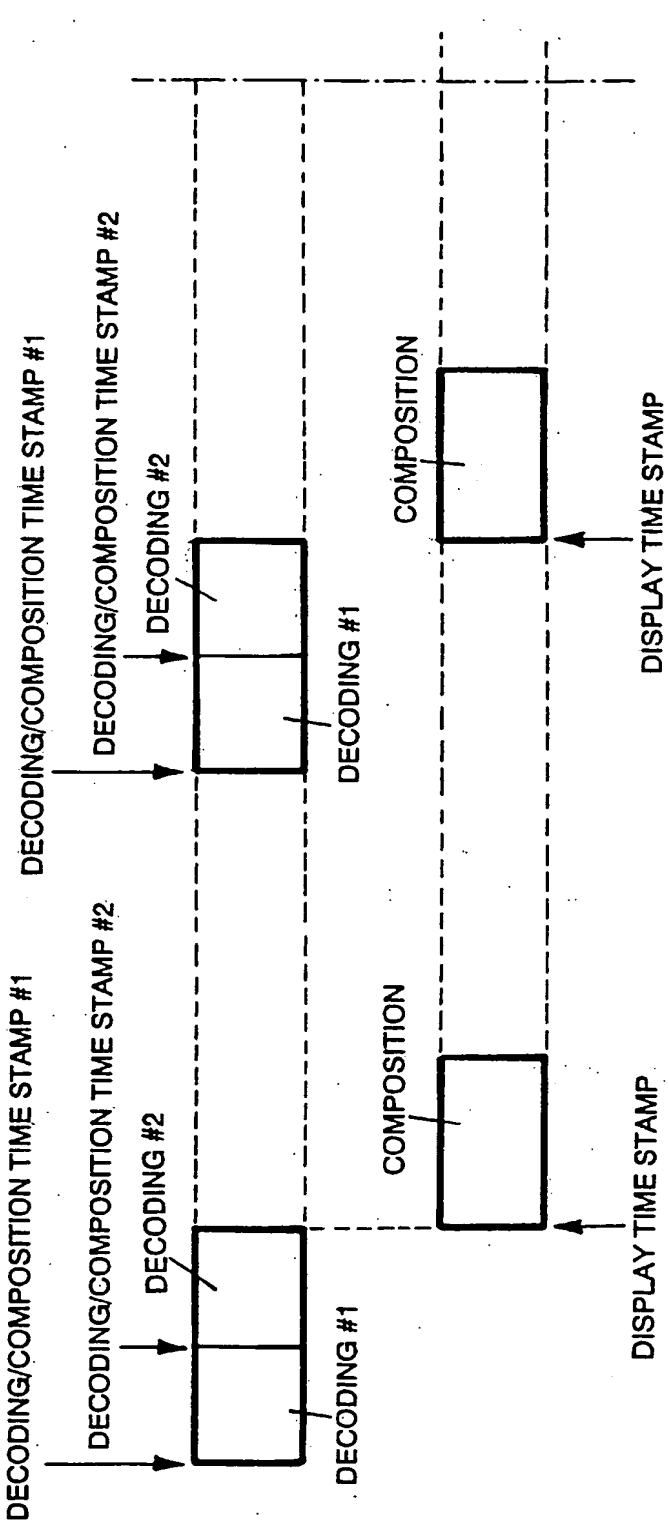
BNSDOCID: <EP_____0924934A1_I_>

FIG.41



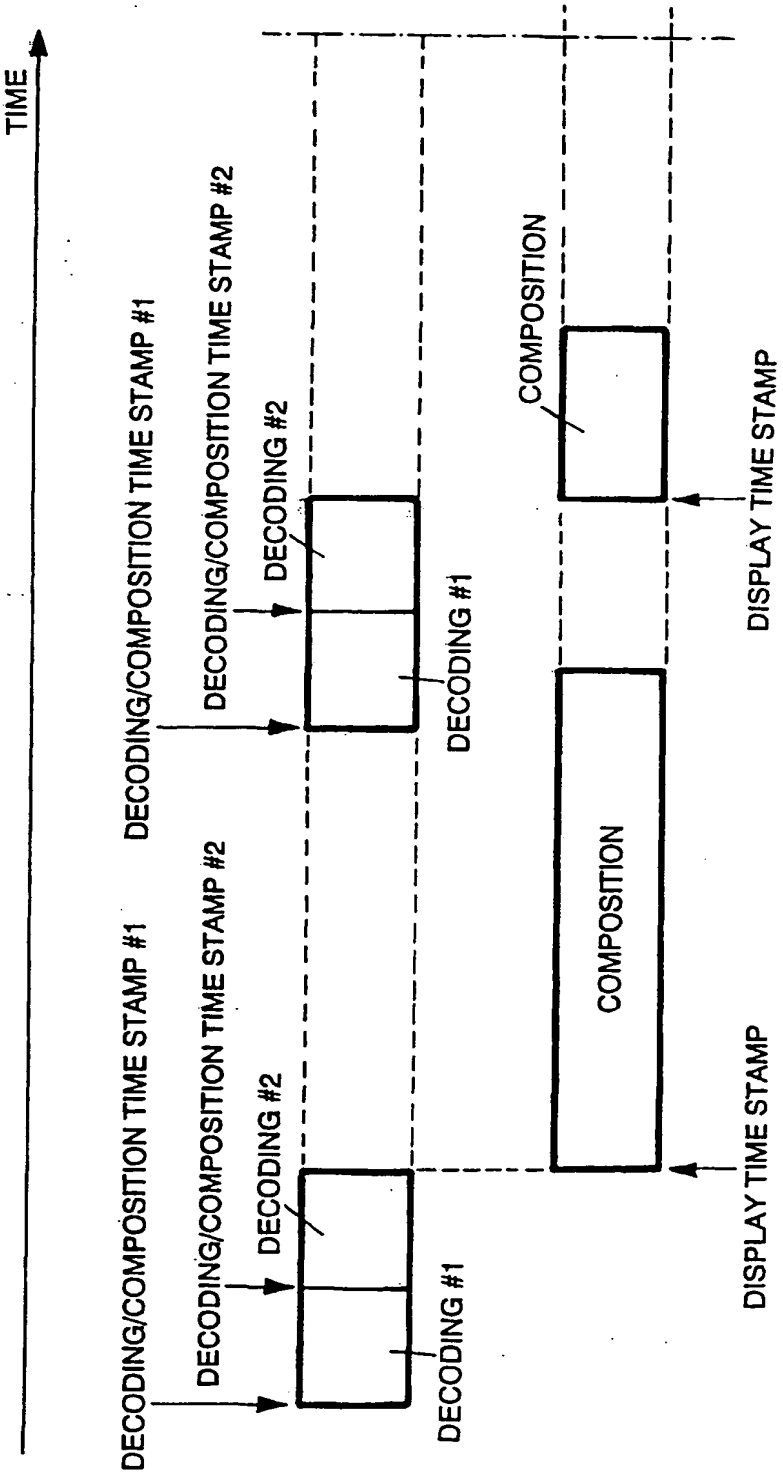
ENSDOCID: <EP_____0924934A1_I_>

FIG.42



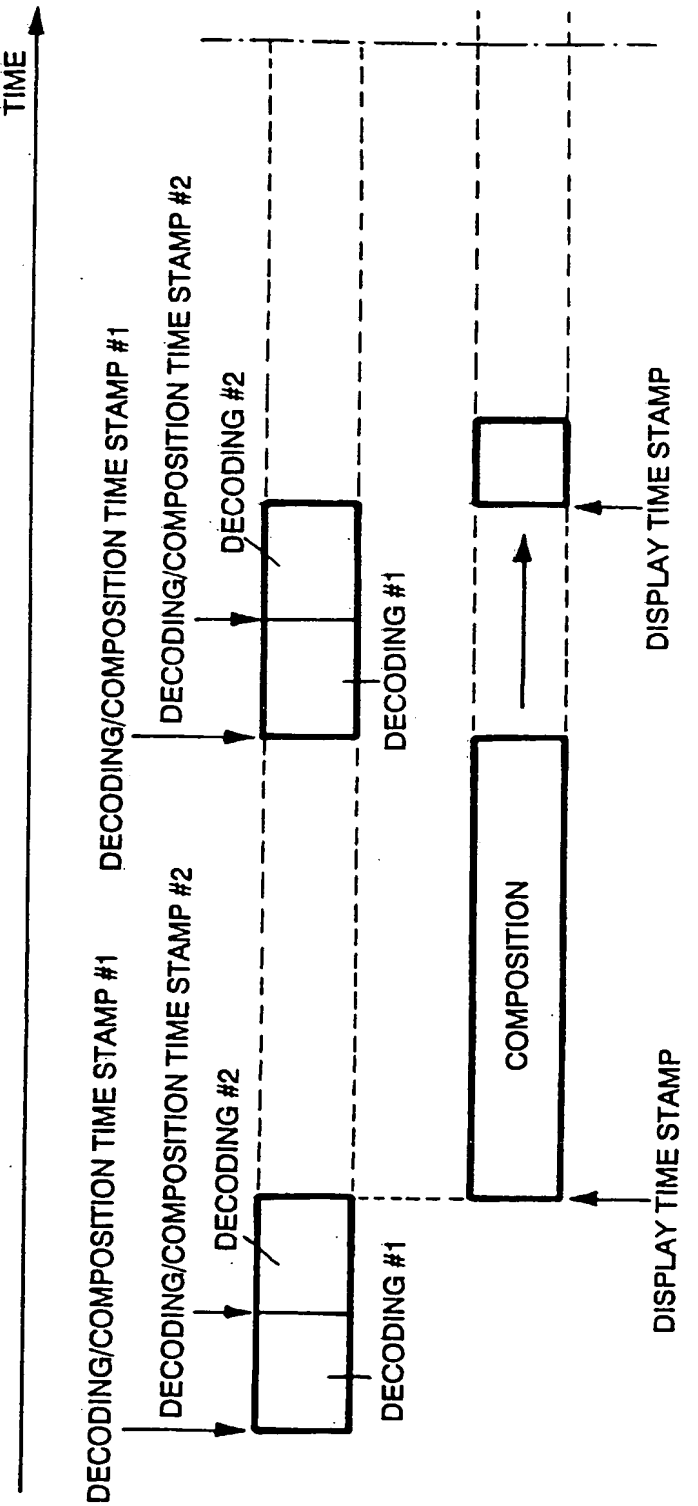
BNSDOCID: <EP_____0924934A1_I_>

FIG.43



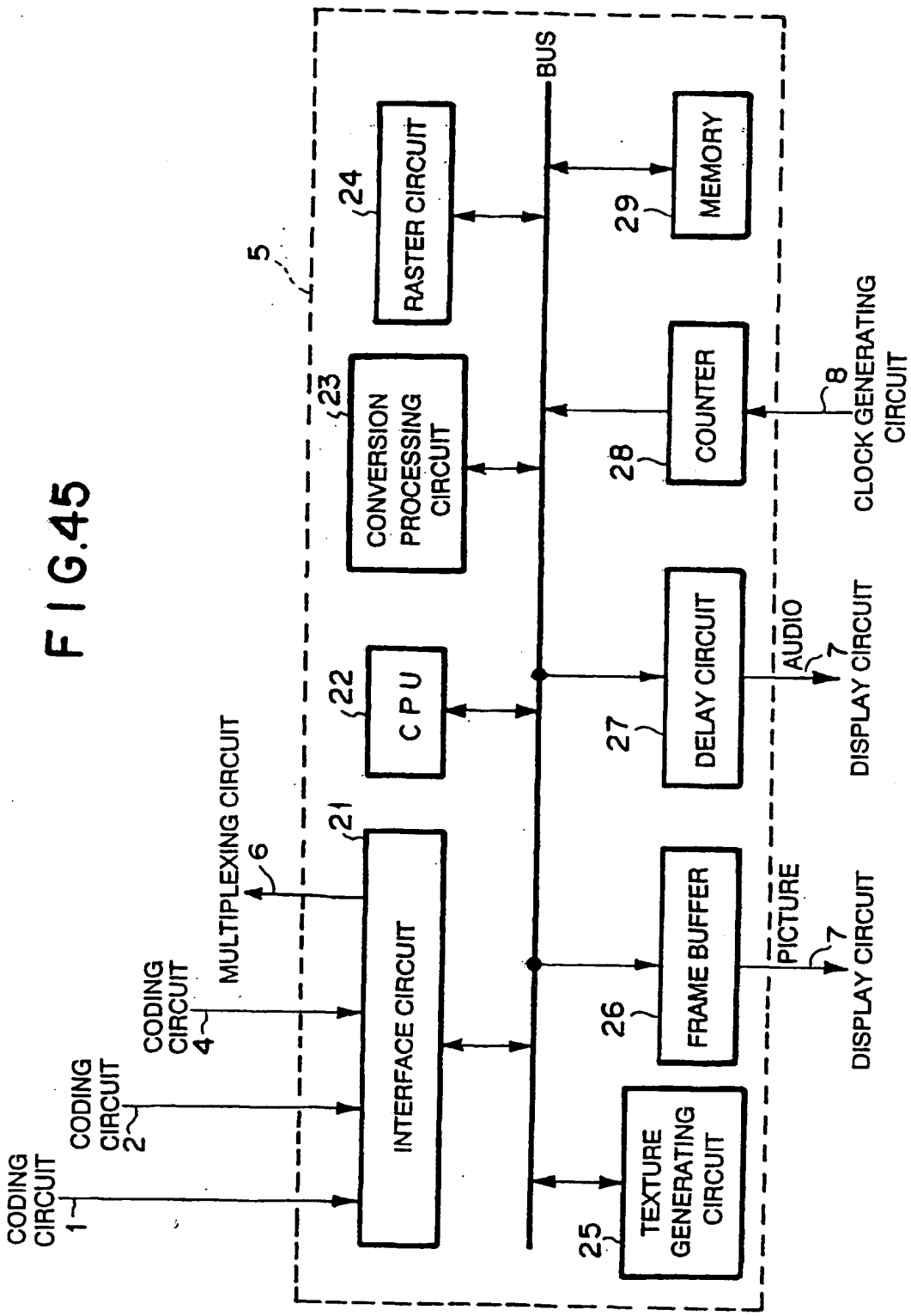
BNSDOCID: <EP_____0924934A1_I_>

FIG.44



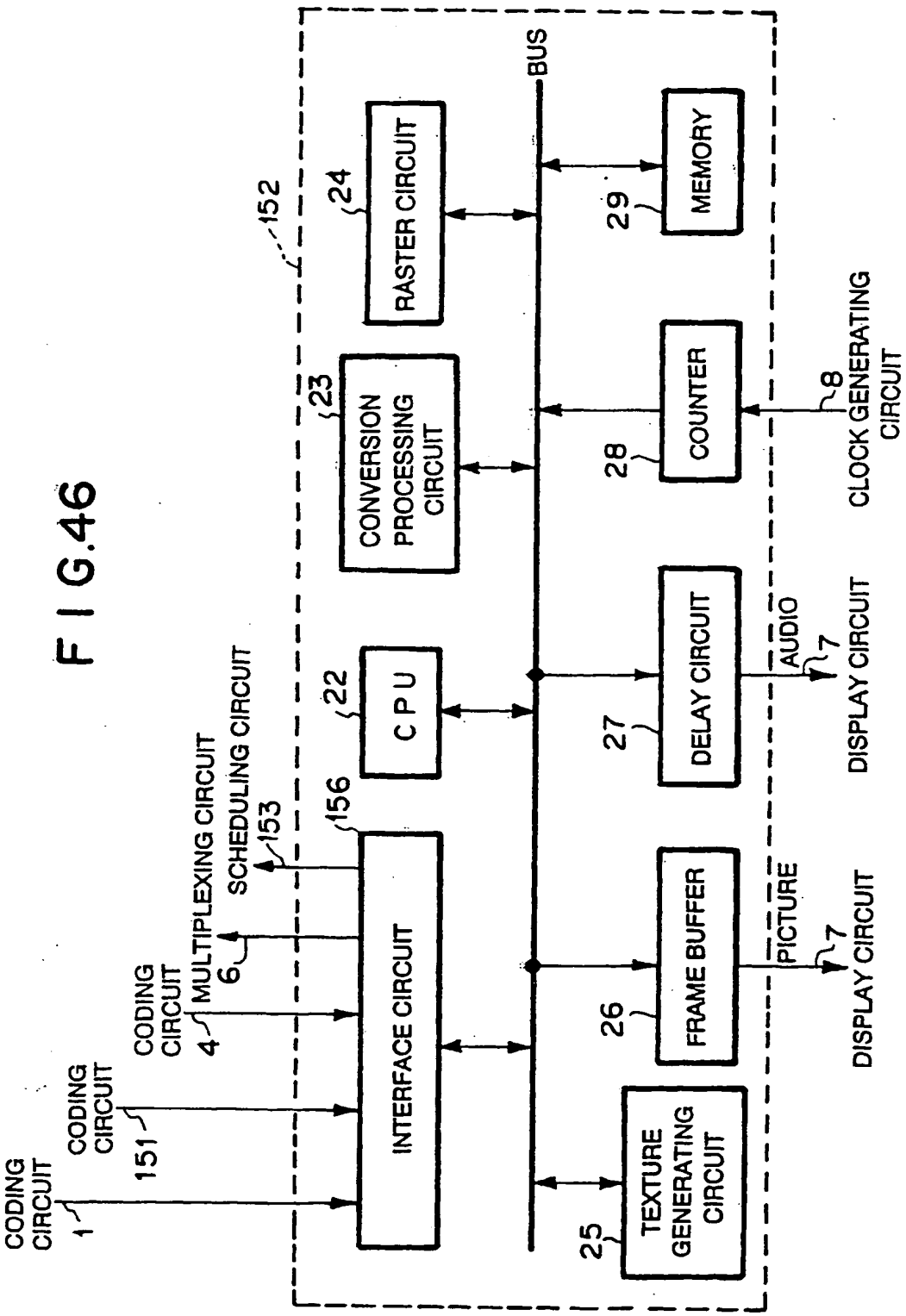
BNSDOCID: <EP_____0924934A1_I_>

FIG.45



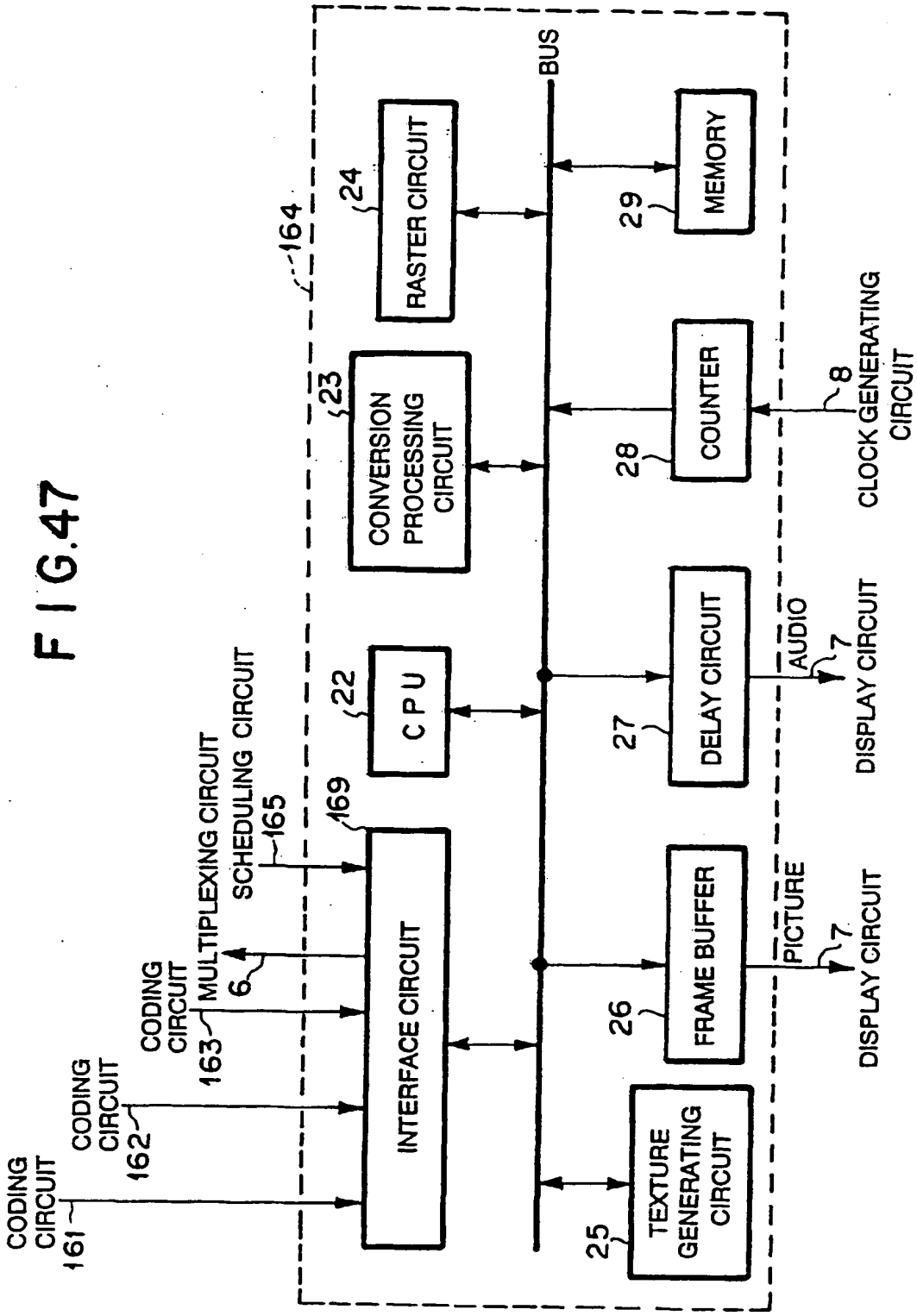
INSDOCID: <EP_____0924934A1_I_>

FIG.46



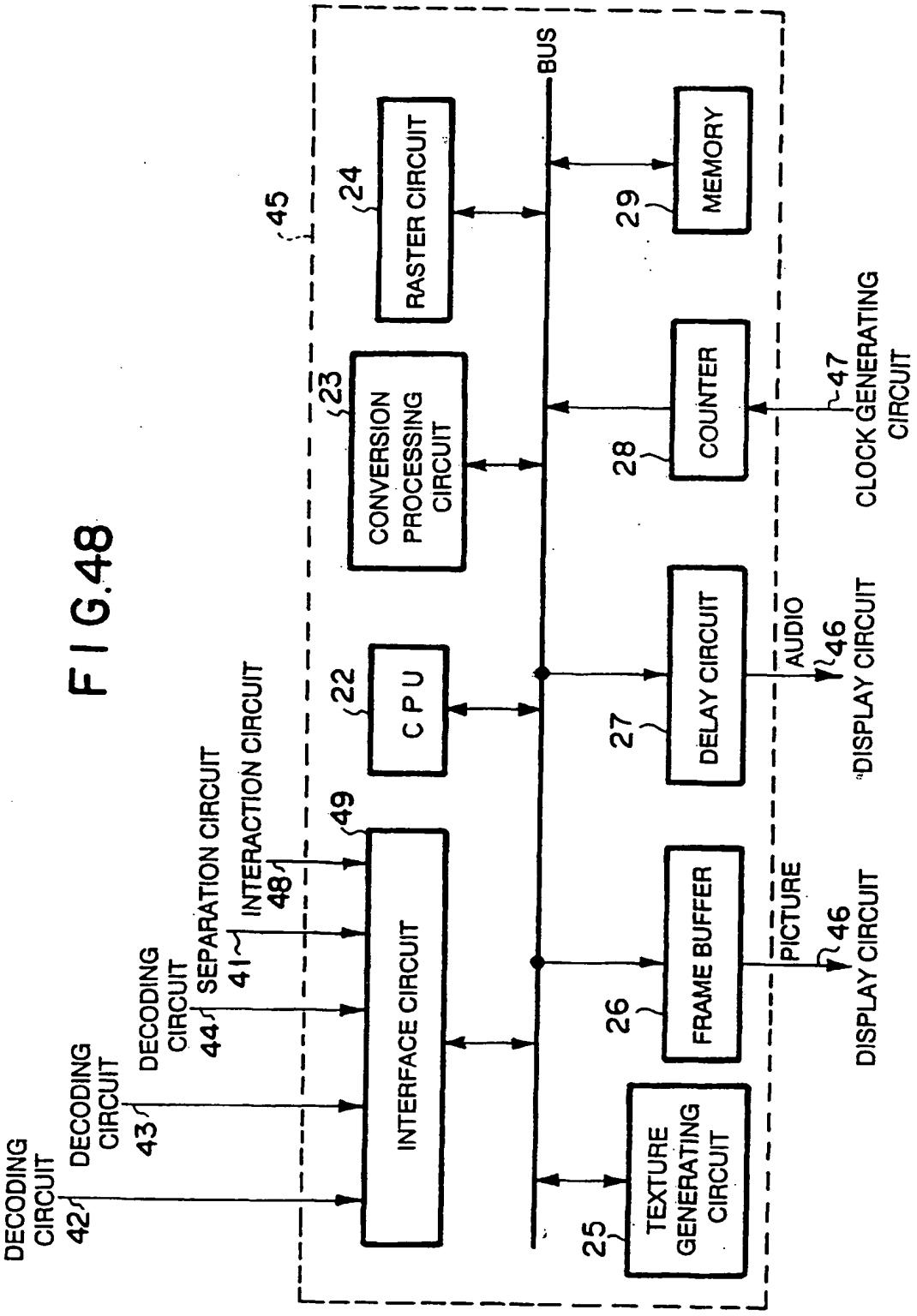
BNSDOCID: <EP_____0924934A1_I_>

FIG.47



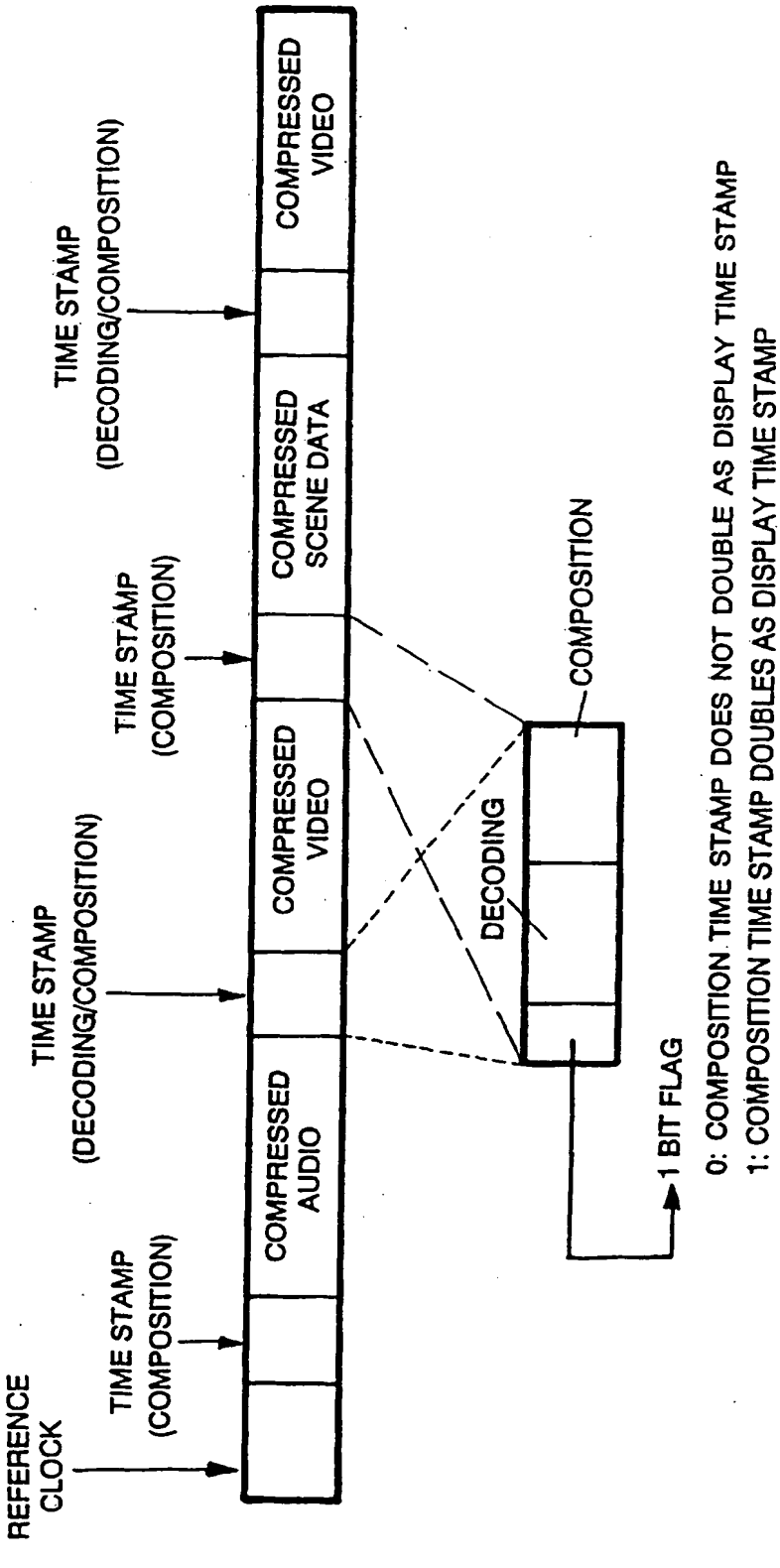
BNSDOCID: <EP_____0924934A1_I_>

FIG.48



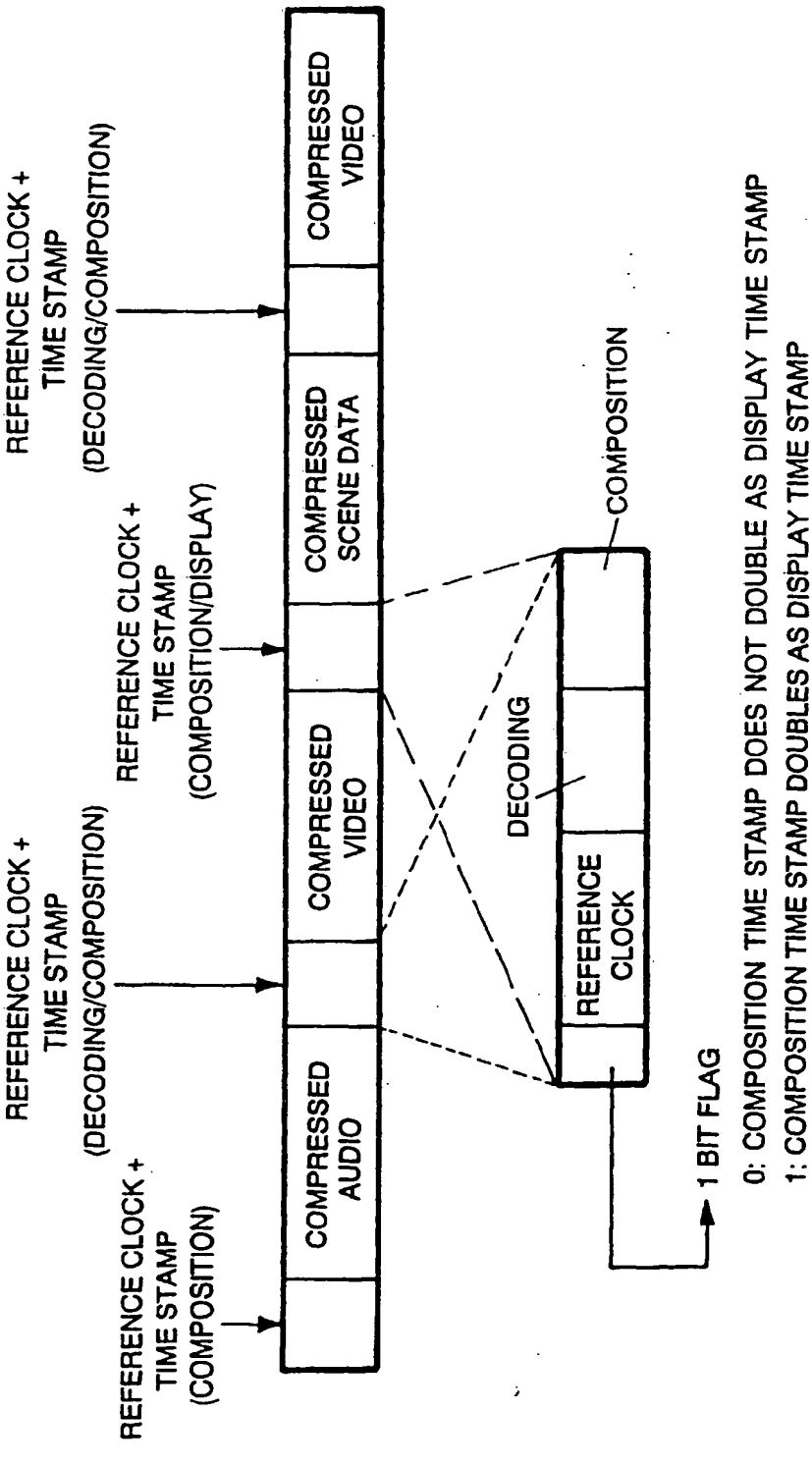
BNSDOCID: <EP_____0924934A1_I_>

FIG.49



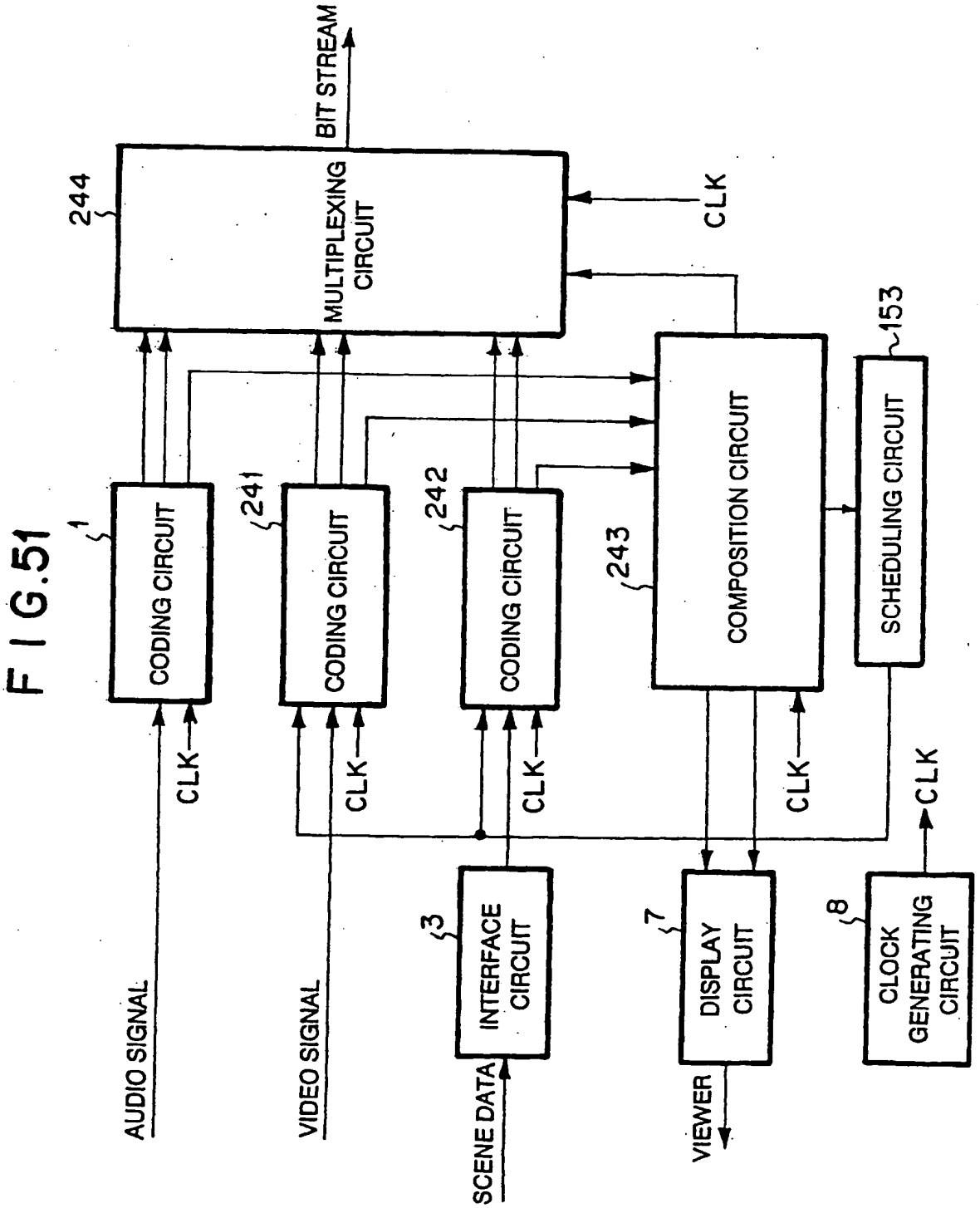
BNSDOCID: <EP_____0924934A1_I_>

FIG.50



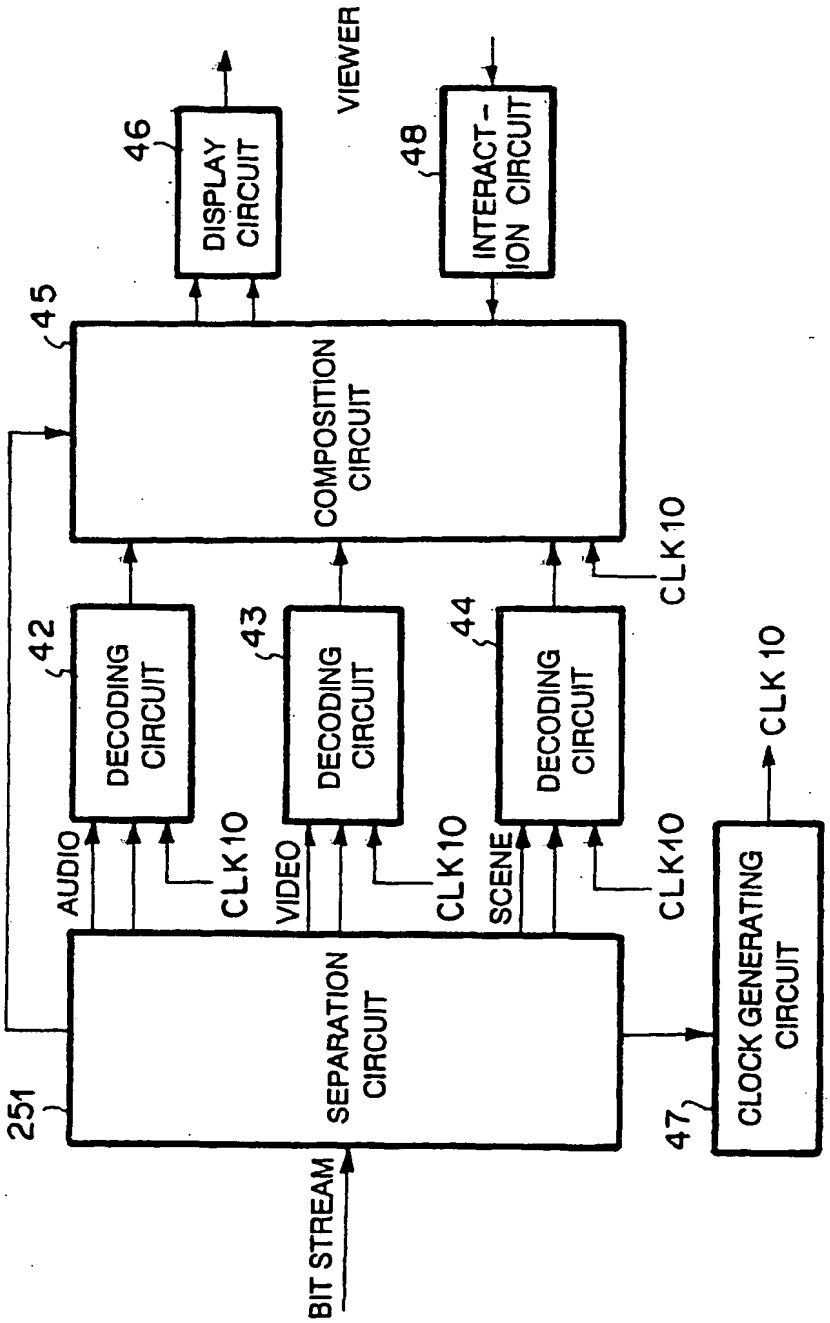
BNSDOCID: <EP_____0924934A1_I_>

FIG.51



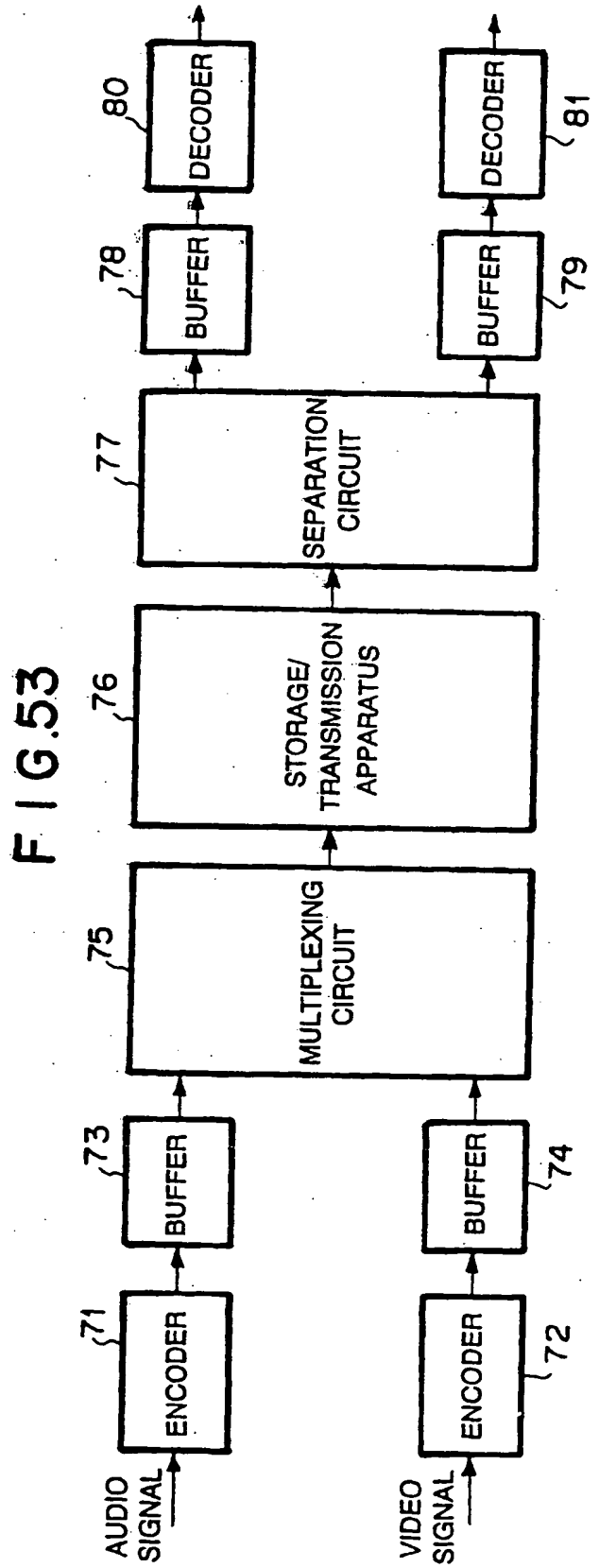
BNSDOCID: <EP_____0924934A1_I_>

FIG.52



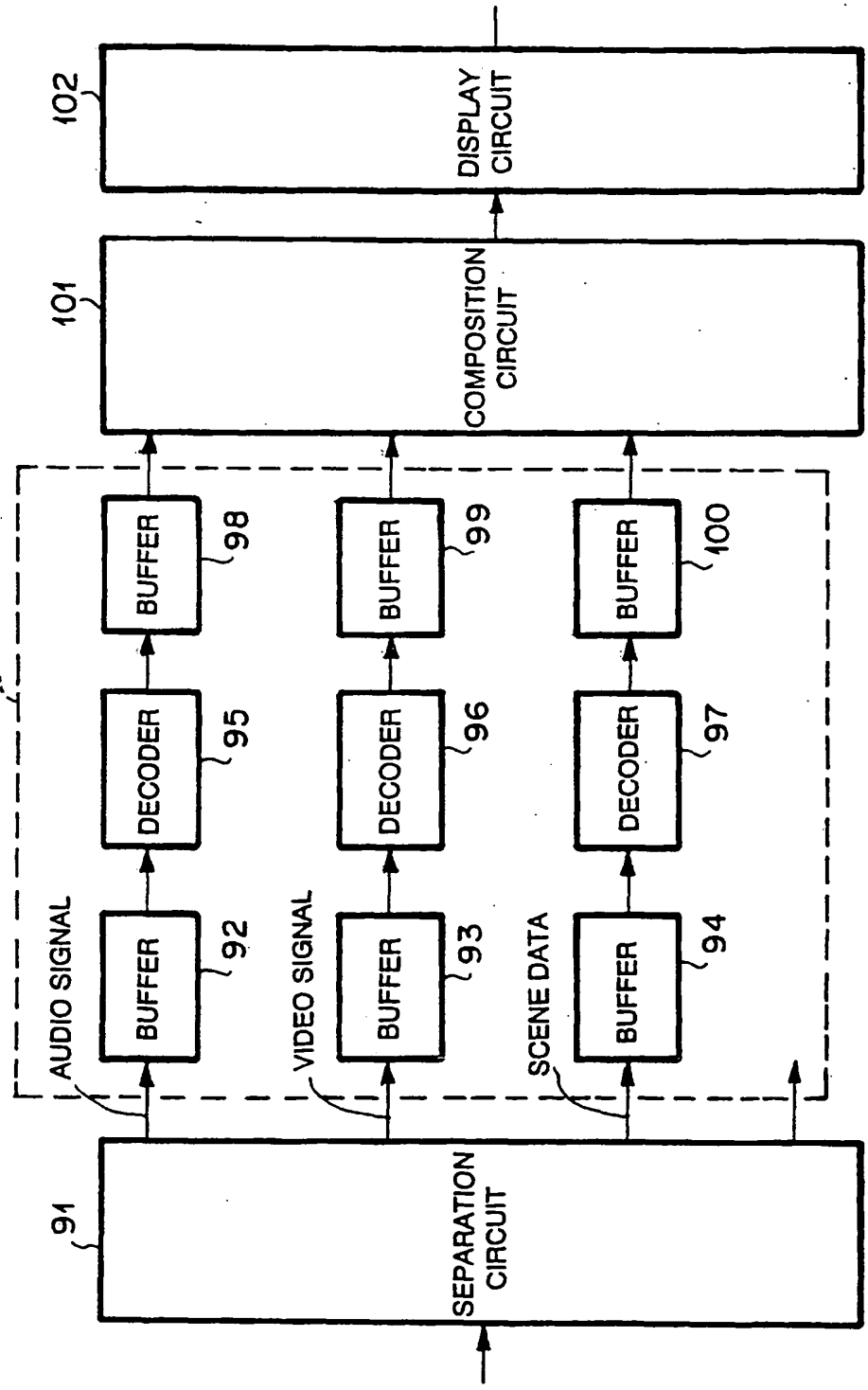
BNSDOCID: <EP_____0924934A1_I_>

FIG.53



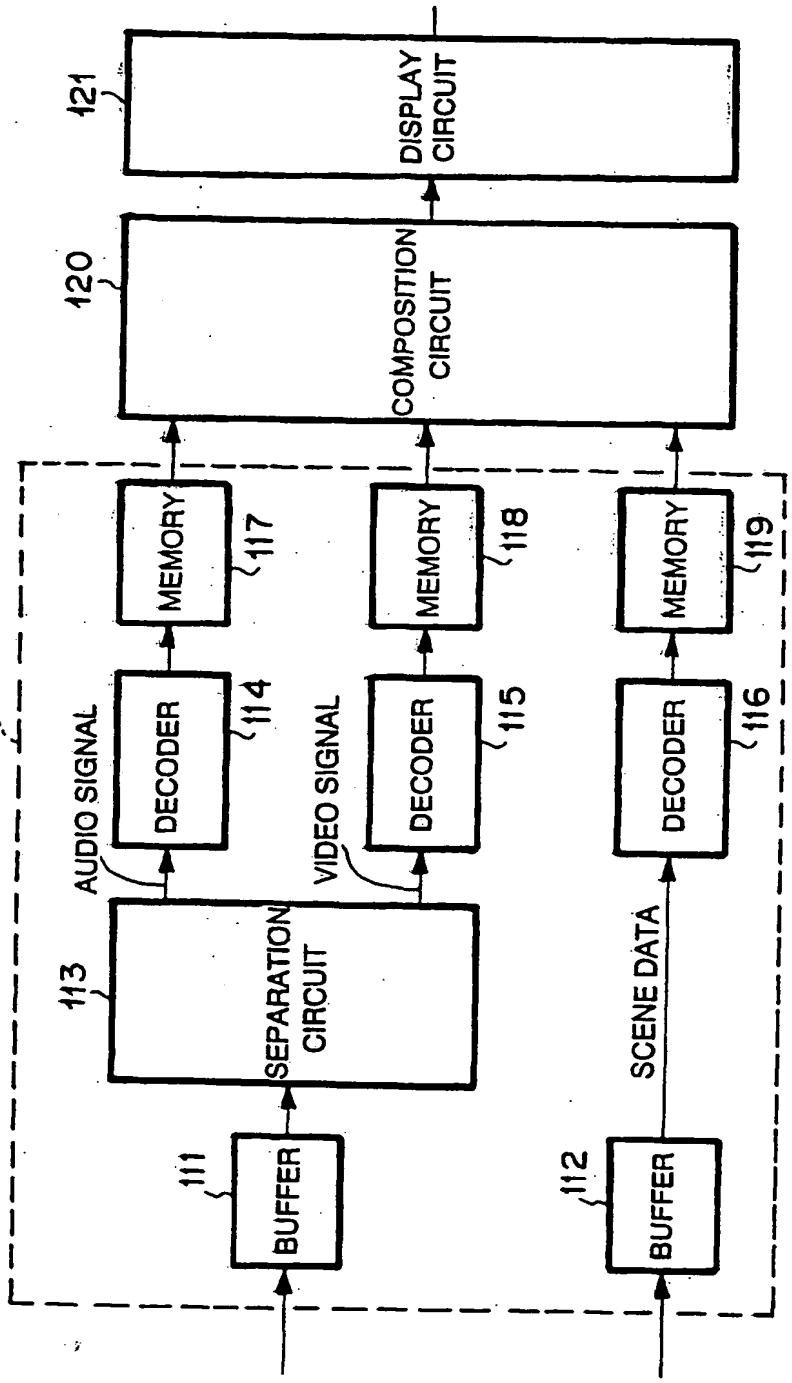
BNSDOCID: <EP_____0924934A1_I_>

FIG.54



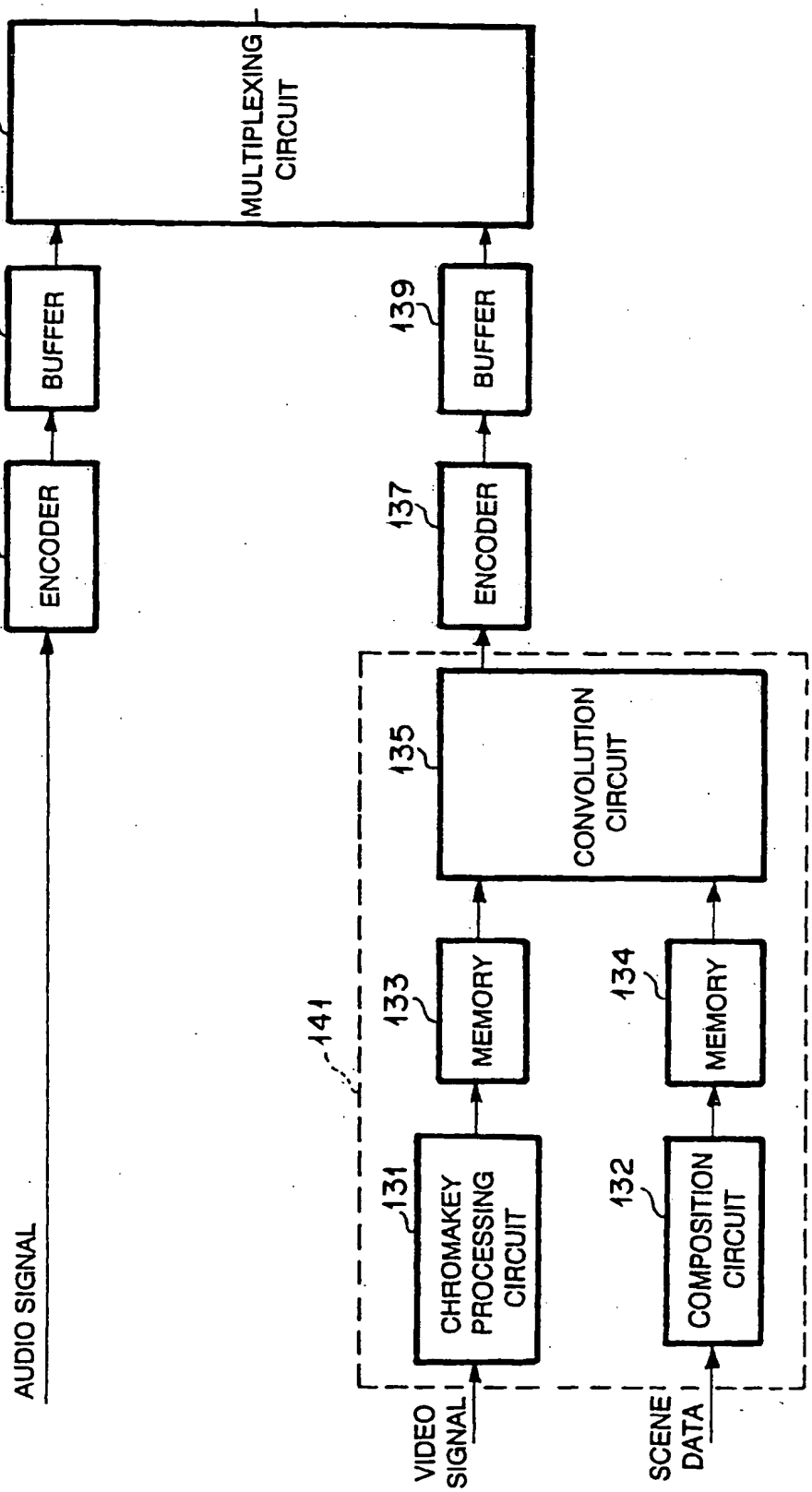
BNSDOCID: <EP_____0924934A1_I_>

FIG.55



BNSDOCID: <EP_____0924934A1_I_>

FIG.56



BNSDOCID: <EP_____0924934A1_I_>

European Patent Office

EUROPEAN SEARCH REPORT

Application Number
EP 98 12 4300

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int.Cl.6)
X	AVARO O ET AL: "The MPEG-4 systems and description languages: A way ahead in audio visual information representation" SIGNAL PROCESSING. IMAGE COMMUNICATION, vol. 9, no. 4, May 1997, page 385-431 XP004075337	6,9,12, 13	H04N7/52
Y	* page 395, paragraph 3; figure 5 *	1	
Y	* page 397, line 11 - line 17 * * page 424, paragraph 2 *	1	
X	DOENGES P K ET AL: "Audio/video and synthetic graphics/audio for mixed media" SIGNAL PROCESSING. IMAGE COMMUNICATION, vol. 9, no. 4, May 1997, page 433-463 XP004075338	6,9,12, 13	
A	* page 443, column 1, paragraph 2 - page 445, column 1, paragraph 1 * * page 456, column 2, paragraph 3 - page 458, column 1, line 2 *	2-5,7,8, 10,11	
Y	WO 97 30551 A (GUEDALIA JACOB LEON ;OLIVR CORP LTD (IL)) 21 August 1997 * page 39, line 10 - page 40, last line; figure 1 *	1	TECHNICAL FIELDS SEARCHED (Int.Cl.6) H04N

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
BERLIN	6 April 1999	Raeymaekers, P

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons

& : member of the same patent family, corresponding document

EPO FORM 1503 03.82 (P04C01)

BNSDOCID: <EP_____0924934A1_I_>

ANNEX TO THE EUROPEAN SEARCH REPORT ON EUROPEAN PATENT APPLICATION NO. EP 98 12 4300

This annex lists the patent family members relating to the patent documents cited in the above-mentioned European search report. The members are as contained in the European Patent Office EDP file on . The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

06-04-1999

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9730551 A	21-08-1997	AU 1616597 A	02-09-1997
		EP 0886968 A	30-12-1998

EPO FORM P0459

For more details about this annex : see Official Journal of the European Patent Office, No. 12/82

BNSDOCID: <EP_____0924934A1_I_>

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
29 March 2001 (29.03.2001)

PCT

(10) International Publication Number
WO 01/22729 A1

(51) International Patent Classification[7]: H04N 5/92

(21) International Application Number: PCT/US00/25847

(22) International Filing Date:
20 September 2000 (20.09.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/154,713	20 September 1999 (20.09.1999)	US
Not furnished	20 September 2000 (20.09.2000)	US

(71) Applicant: TIVO, INC. [US/US]; 2160 Gold Street, P.O. Box 2160, Alviso, CA 95002 (US).

(72) Inventors: BARTON, James, M.; 101 Sund Avenue, Los Gatos, CA 95032 (US). SMITH, Kevin; 1164 Karen Way, Mountain View, CA 94040 (US). CHAMBERLIN, David; 206 Flynn Avenue, Mountain View, CA 94043 (US). LOOK, Howard; 576 Palo Alto Avenue, Mountain View, CA 94041 (US).

(74) Agents: GLENN, Michael et al.; Glenn Patent Group, 3475 Edison Way, Ste. L., Menlo Park, CA 94025 (US).

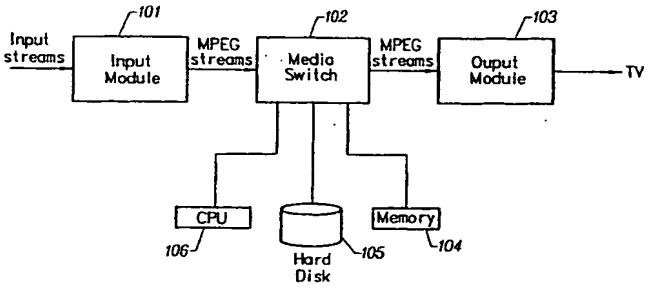
(81) Designated States *(national)*: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *With international search report.*
— *Before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments.*

[Continued on next page]

(54) Title: CLOSED CAPTION TAGGING SYSTEM



(57) Abstract: A closed caption tagging system provides a mechanism for inserting tags into an audio or video television broadcast stream prior to or at the time of transmission. The tags contain command and control information that the receiver translates and acts upon. The receiver receives the broadcast stream and detects and processes the tags within the broadcast stream which is stored on a storage device that resides on the receiver. Program material from the broadcast stream is played back to the viewer from the storage device. The receiver performs the appropriate actions in response to the tags. Tags indicate the start and end points of a program segment. The receiver skips over a program segment during playback in response to the viewer pressing a button on a remote input device or it automatically skips over program segments depending on the viewer's preferences. Program segments such as commercials are automatically replaced by the receiver with new program segments that are selected based on various criteria. Menus, icons, and Web pages are displayed to the viewer based on information included in a tag. The viewer interacts with the menu, icon, or Web page through an input device with the receiver performing the associated actions. If a menu or actions requires that the viewer exits from the playback of the program material, then the receiver saves the exit point and returns the viewer back to the same exit point when the viewer has completed the interaction session. Menus and icons are used to generate leads, generate sales, and schedule the recording of programs. A one-touch recording option is provided. An icon is displayed to the viewer telling the viewer that an advertised program is available for recording at a future time. The viewer presses a single button on an input device causing the receiver to schedule the program for recording.

WO 01/22729 A1

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

CLOSED CAPTION TAGGING SYSTEM

BACKGROUND OF THE INVENTION

TECHNICAL FIELD

The invention relates to the processing of multimedia audio and video streams. More particularly, the invention relates to the tagging of multimedia audio and video television streams.

DESCRIPTION OF THE PRIOR ART

The Video Cassette Recorder (VCR) has changed the lives of television (TV) viewers throughout the world. The VCR has offered viewers the flexibility to time-shift TV programs to match their lifestyles.

The viewer stores TV programs onto magnetic tape using the VCR. The VCR gives the viewer the ability to play, rewind, fast forward and pause the stored program material. These functions enable the viewer to pause the program playback whenever he desires, fast forward through unwanted program material or commercials, and to replay favorite scenes. However, a VCR cannot both capture and play back information at the same time.

Digital Video Recorders (DVR) have recently entered into the marketplace. DVRs allow the viewer to store TV programs on a hard disk. This has freed the viewer from the magnetic tape realm. Viewers can pause, rewind, and fast forward live broadcast programs. However, the functionality of DVRs extends beyond recording programs.

Having programs stored locally in a digital form gives the programmer many more options than were previously available. Advertisements (commercials) can now be dynamically replaced and specifically targeted to the particular viewer based on his or her viewing habits. The commercials can be stored locally on the viewer's DVR and shown at any time.

DVRs allow interactive programming with the viewer. Generally, promotions for future shows are displayed to viewers during the normal broadcast programs. Viewers must then remember the date, time, and channel that the program will be aired on to record or view the program. DVRs allow the viewer to schedule the recording of the program immediately.

The only drawback is that the current generation of DVRs do not have the capability to interact with the viewer at this level. There is no means by which to notify the DVR that commercials are directly tied to a certain program or other advertisements. Further, there is no way to tell the DVR that a commercial can be replaced.

It would be advantageous to provide a closed caption tagging system that gives the content provider the ability to send frame specific data across broadcast media. It would further be advantageous to provide a closed caption tagging system that allows the receiver to dynamically interact with the viewer and configure itself based on program content.

SUMMARY OF THE INVENTION

The invention provides a closed caption tagging system. The invention allows content providers to send frame specific data and commands integrated into video and audio television streams across broadcast media. In addition, the invention allows the receiver to dynamically interact with the viewer and configure itself based on video and audio stream content.

A preferred embodiment of the invention provides a mechanism for inserting tags into an audio or video television broadcast stream. Tags are inserted into the broadcast stream prior to or at the time of transmission. The tags contain command and control information that the receiver translates and acts upon.

The receiver receives the broadcast stream and detects and processes the tags within the broadcast stream. The broadcast stream is stored on a storage device that resides on the receiver. Program material from the broadcast stream is played back to the viewer from the storage device.

During the tag processing stage, the receiver performs the appropriate actions in

INSDOCID: <WO_____0122729A1_I_>

response to the tags. The tags offer a great amount of flexibility to the content provider or system administrator to create a limitless amount of operations.

Tags indicate the start and end points of a program segment. The receiver skips over a program segment during playback in response to the viewer pressing a button on a remote input device. The receiver also automatically skips over program segments depending on the viewer's preferences.

Program segments such as commercials are automatically replaced by the receiver with new program segments. New program segments are selected based on various criteria such as the locale, time of day, program material, viewer's viewing habits, viewer's program preferences, or the viewer's personal information. The new program segments are stored remotely or locally on the receiver.

Menus, icons, and Web pages are displayed to the viewer based on information included in a tag. The viewer interacts with the menu, icon, or Web page through an input device. The receiver performs the actions associated with the menu, icon, or Web page and the viewer's input. If a menu or action requires that the viewer exit from the playback of the program material, then the receiver saves the exit point and returns the viewer back to the same exit point when the viewer has completed the interaction session.

Menus and icons are used to generate leads, generate sales, and schedule the recording of programs. A one-touch recording option is provided. An icon is displayed to the viewer telling the viewer that an advertised program is available for recording at a future time. The viewer presses a single button on an input device causing the receiver to schedule the program for recording. The receiver will also record the current program in the broadcast stream onto the storage device based on information included in a tag.

Tags are used to create indexes in program material. This allows the viewer to jump to particular indexes in a program.

Other aspects and advantages of the invention will become apparent from the following detailed description in combination with the accompanying drawings, illustrating, by way of example, the principles of the invention.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 is a block schematic diagram of a high level view of a preferred embodiment of the invention according to the invention;

Fig. 2 is a block schematic diagram of a preferred embodiment of the invention using multiple input and output modules according to the invention;

Fig. 3 is a schematic diagram of an Moving Pictures Experts Group (MPEG) data stream and its video and audio components according to the invention;

Fig. 4 is a block schematic diagram of a parser and four direct memory access (DMA) input engines contained in the Media Switch according to the invention;

Fig. 5 is a schematic diagram of the components of a packetized elementary stream (PES) buffer according to the invention;

Fig. 6 is a schematic diagram of the construction of a PES buffer from the parsed components in the Media Switch output circular buffers;

Fig. 7 is a block schematic diagram of the Media Switch and the various components that it communicates with according to the invention;

Fig. 8 is a block schematic diagram of a high level view of the program logic according to the invention;

Fig. 9 is a block schematic diagram of a class hierarchy of the program logic according to the invention;

Fig. 10 is a block schematic diagram of a preferred embodiment of the clip cache component of the invention according to the invention;

Fig. 11 is a block schematic diagram of a preferred embodiment of the invention that emulates a broadcast studio video mixer according to the invention;

Fig. 12 is a block schematic diagram of a closed caption parser according to the invention;

Fig. 13 is a block schematic diagram of a high level view of a preferred embodiment of the invention utilizing a VCR as an integral component of the invention according to the invention;

Fig. 14 is a block schematic diagram of a preferred embodiment of the invention for inserting tags into a video stream according to the invention;

Fig. 15 is a block schematic diagram of a server-based preferred embodiment of the invention for inserting tags into a video stream according to the invention;

Fig. 16 is a diagram of a user interface for inserting tags into a video stream according to the invention;

Fig. 17 is a diagram of a screen with an alert icon displayed in the lower left corner of the screen according to the invention;

Fig. 18 is a block schematic diagram of the transmission route of a video stream according to the invention;

Fig. 19 is a block schematic diagram of the tagging of the start and end of a program segment of a video stream and the playback of a new program segment according to the invention;

Fig. 20 is a block schematic diagram of a preferred embodiment of the invention that interprets tags inserted into a video stream according to the invention;

Fig. 21 is a diagram of a screen displaying program recording options according to the invention;

Fig. 22 is a diagram of a viewer remote control device according to the invention; and

Fig. 23 is a block schematic diagram of a series of screens for lead and sale generation according to the invention.

DETAILED DESCRIPTION OF THE INVENTION

The invention is embodied in a closed caption tagging system. A system according to the invention allows content providers to send frame specific data and commands integrated into video and audio television streams across broadcast media. The invention additionally allows the receiver to dynamically interact with the viewer and configure itself based on video and audio stream content.

A preferred embodiment of the invention provides a tagging and interpretation system that allows a content provider to tag, in a frame specific manner, video and audio streams transmitted over television broadcast media. A receiver interprets and acts upon the tags embedded in the received stream. The tag data allow the receiver to dynamically interact with the viewer through menus and action icons. The tags also provide for the dynamic configuration of the receiver.

Referring to Fig. 1, a preferred embodiment of the invention has an Input Section 101, Media Switch 102, and an Output Section 103. The Input Section 101 takes television (TV) input streams in a multitude of forms, for example, National Television Standards Committee (NTSC) or PAL broadcast, and digital forms such as Digital Satellite System (DSS), Digital Broadcast Services (DBS), or Advanced Television Standards Committee (ATSC). DBS, DSS and ATSC are based on standards called Moving Pictures Experts Group 2 (MPEG2) and MPEG2 Transport. MPEG2 Transport is a standard for formatting the digital data stream from the TV source transmitter so that a TV receiver can disassemble the input stream to find programs in the multiplexed signal. The Input Section 101 produces MPEG streams. An MPEG2 transport multiplex supports multiple programs in the same broadcast channel, with multiple video and audio feeds and private data. The Input Section 101 tunes the channel to a particular program, extracts a specific MPEG program out of it, and feeds it to the rest of the system. Analog TV signals are encoded into a similar MPEG format using separate video and audio encoders, such that the remainder of the system is unaware of how the signal was obtained. Information may be modulated into the Vertical Blanking Interval (VBI) of the analog TV signal in a number of standard ways; for example, the North American Broadcast Teletext Standard (NABTS) may be used to modulate information onto lines 10 through 20 of an NTSC

3NSDOCID: <WO_____0122729A1_I_>

signal, while the FCC mandates the use of line 21 for Closed Caption (CC) and Extended Data Services (EDS). Such signals are decoded by the input section and passed to the other sections as if they were delivered via an MPEG2 private data channel.

The Media Switch 102 mediates between a microprocessor CPU 106, hard disk or storage device 105, and memory 104. Input streams are converted to an MPEG stream and sent to the Media Switch 102. The Media Switch 102 buffers the MPEG stream into memory. It then performs two operations if the user is watching real time TV: the stream is sent to the Output Section 103 and it is written simultaneously to the hard disk or storage device 105.

The Output Section 103 takes MPEG streams as input and produces an analog TV signal according to the NTSC, PAL, or other required TV standards. The Output Section 103 contains an MPEG decoder, On-Screen Display (OSD) generator, analog TV encoder and audio logic. The OSD generator allows the program logic to supply images which will be overlayed on top of the resulting analog TV signal. Additionally, the Output Section can modulate information supplied by the program logic onto the VBI of the output signal in a number of standard formats, including NABTS, CC and EDS.

With respect to Fig. 2, the invention easily expands to accommodate multiple Input Sections (tuners) 201, 202, 203, 204, each can be tuned to different types of input. Multiple Output Modules (decoders) 206, 207, 208, 209 are added as well. Special effects such as picture in a picture can be implemented with multiple decoders. The Media Switch 205 records one program while the user is watching another. This means that a stream can be extracted off the disk while another stream is being stored onto the disk.

Referring to Fig. 3, the incoming MPEG stream 301 has interleaved video 302, 305, 306 and audio 303, 304, 307 segments. These elements must be separated and recombined to create separate video 308 and audio 309 streams or buffers. This is necessary because separate decoders are used to convert MPEG elements back into audio or video analog components. Such separate delivery requires that time sequence information be generated so that the decoders may be properly synchronized for accurate playback of the signal.

The Media Switch enables the program logic to associate proper time sequence

information with each segment, possibly embedding it directly into the stream. The time sequence information for each segment is called a time stamp. These time stamps are monotonically increasing and start at zero each time the system boots up. This allows the invention to find any particular spot in any particular video segment. For example, if the system needs to read five seconds into an incoming contiguous video stream that is being cached, the system simply has to start reading forward into the stream and look for the appropriate time stamp.

A binary search can be performed on a stored file to index into a stream. Each stream is stored as a sequence of fixed-size segments enabling fast binary searches because of the uniform timestamping. If the user wants to start in the middle of the program, the system performs a binary search of the stored segments until it finds the appropriate spot, obtaining the desired results with a minimal amount of information. If the signal were instead stored as an MPEG stream, it would be necessary to linearly parse the stream from the beginning to find the desired location.

With respect to Fig. 4, the Media Switch contains four input Direct Memory Access (DMA) engines 402, 403, 404, 405 each DMA engine has an associated buffer 410, 411, 412, 413. Conceptually, each DMA engine has a pointer 406, a limit for that pointer 407, a next pointer 408, and a limit for the next pointer 409. Each DMA engine is dedicated to a particular type of information, for example, video 402, audio 403, and parsed events 405. The buffers 410, 411, 412, 413 are circular and collect the specific information. The DMA engine increments the pointer 406 into the associated buffer until it reaches the limit 407 and then loads the next pointer 408 and limit 409. Setting the pointer 406 and next pointer 408 to the same value, along with the corresponding limit value creates a circular buffer. The next pointer 408 can be set to a different address to provide vector DMA.

The input stream flows through a parser 401. The parser 401 parses the stream looking for MPEG distinguished events indicating the start of video, audio or private data segments. For example, when the parser 401 finds a video event, it directs the stream to the video DMA engine 402. The parser 401 buffers up data and DMAs it into the video buffer 410 through the video DMA engine 402. At the same time, the parser 401 directs an event to the event DMA engine 405 which generates an event into the event buffer 413. When the parser 401 sees an audio event, it redirects the byte stream to the audio DMA engine 403 and

generates an event into the event buffer 413. Similarly, when the parser 401 sees a private data event, it directs the byte stream to the private data DMA engine 404 and directs an event to the event buffer 413. The Media Switch notifies the program logic via an interrupt mechanism when events are placed in the event buffer.

Referring to Figs. 4 and 5, the event buffer 413 is filled by the parser 401 with events. Each event 501 in the event buffer has an offset 502, event type 503, and time stamp field 504. The parser 401 provides the type and offset of each event as it is placed into the buffer. For example, when an audio event occurs, the event type field is set to an audio event and the offset indicates the location in the audio buffer 411. The program logic knows where the audio buffer 411 starts and adds the offset to find the event in the stream. The address offset 502 tells the program logic where the next event occurred, but not where it ended. The previous event is cached so the end of the current event can be found as well as the length of the segment.

With respect to Figs. 5 and 6, the program logic reads accumulated events in the event buffer 602 when it is interrupted by the Media Switch 601. From these events the program logic generates a sequence of logical segments 603 which correspond to the parsed MPEG segments 615. The program logic converts the offset 502 into the actual address 610 of each segment, and records the event length 609 using the last cached event. If the stream was produced by encoding an analog signal, it will not contain Program Time Stamp (PTS) values, which are used by the decoders to properly present the resulting output. Thus, the program logic uses the generated time stamp 504 to calculate a simulated PTS for each segment and places that into the logical segment timestamp 607. In the case of a digital TV stream, PTS values are already encoded in the stream. The program logic extracts this information and places it in the logical segment timestamp 607.

The program logic continues collecting logical segments 603 until it reaches the fixed buffer size. When this occurs, the program logic generates a new buffer, called a Packetized Elementary Stream (PES) 605 buffer containing these logical segments 603 in order, plus ancillary control information. Each logical segment points 604 directly to the circular buffer, *e.g.*, the video buffer 613, filled by the Media Switch 601. This new buffer is then passed to other logic components, which may further process the stream in the buffer in some way, such as

presenting it for decoding or writing it to the storage media. Thus, the MPEG data is not copied from one location in memory to another by the processor. This results in a more cost effective design since lower memory bandwidth and processor bandwidth is required.

A unique feature of the MPEG stream transformation into PES buffers is that the data associated with logical segments need not be present in the buffer itself, as presented above. When a PES buffer is written to storage, these logical segments are written to the storage medium in the logical order in which they appear. This has the effect of gathering components of the stream, whether they be in the video, audio or private data circular buffers, into a single linear buffer of stream data on the storage medium. The buffer is read back from the storage medium with a single transfer from the storage media, and the logical segment information is updated to correspond with the actual locations in the buffer 606. Higher level program logic is unaware of this transformation, since it handles only the logical segments, thus stream data is easily managed without requiring that the data ever be copied between locations in DRAM by the CPU.

A unique aspect of the Media Switch is the ability to handle high data rates effectively and inexpensively. It performs the functions of taking video and audio data in, sending video and audio data out, sending video and audio data to disk, and extracting video and audio data from the disk on a low cost platform. Generally, the Media Switch runs asynchronously and autonomously with the microprocessor CPU, using its DMA capabilities to move large quantities of information with minimal intervention by the CPU.

Referring to Fig. 7, the input side of the Media Switch 701 is connected to an MPEG encoder 703. There are also circuits specific to MPEG audio 704 and vertical blanking interval (VBI) data 702 feeding into the Media Switch 701. If a digital TV signal is being processed instead, the MPEG encoder 703 is replaced with an MPEG2 Transport Demultiplexor, and the MPEG audio encoder 704 and VBI decoder 702 are deleted. The demultiplexor multiplexes the extracted audio, video and private data channel streams through the video input Media Switch port.

The parser 705 parses the input data stream from the MPEG encoder 703, audio encoder 704 and VBI decoder 702, or from the transport demultiplexor in the case of a digital TV stream. The parser 705 detects the beginning of all of the

important events in a video or audio stream, the start of all of the frames, the start of sequence headers - all of the pieces of information that the program logic needs to know about in order to both properly play back and perform special effects on the stream, *e.g.* fast forward, reverse, play, pause, fast/slow play, indexing, and fast/slow reverse play.

The parser 705 places tags 707 into the FIFO 706 when it identifies video or audio segments, or is given private data. The DMA 709 controls when these tags are taken out. The tags 707 and the DMA addresses of the segments are placed into the event queue 708. The frame type information, whether it is a start of a video I-frame, video B-frame, video P-frame, video PES, audio PES, a sequence header, an audio frame, or private data packet, is placed into the event queue 708 along with the offset in the related circular buffer where the piece of information was placed. The program logic operating in the CPU 713 examines events in the circular buffer after it is transferred to the DRAM 714.

The Media Switch 701 has a data bus 711 that connects to the CPU 713 and DRAM 714. An address bus 712 is also shared between the Media Switch 701, CPU 713, and DRAM 714. A hard disk or storage device 710 is connected to one of the ports of the Media Switch 701. The Media Switch 701 outputs streams to an MPEG video decoder 715 and a separate audio decoder 717. The audio decoder 717 signals contain audio cues generated by the system in response to the user's commands on a remote control or other internal events. The decoded audio output from the MPEG decoder is digitally mixed 718 with the separate audio signal. The resulting signals contain video, audio, and on-screen displays and are sent to the TV 716.

The Media Switch 701 takes in 8-bit data and sends it to the disk, while at the same time extracts another stream of data off of the disk and sends it to the MPEG decoder 715. All of the DMA engines described above can be working at the same time. The Media Switch 701 can be implemented in hardware using a Field Programmable Gate Array (FPGA), ASIC, or discrete logic.

Rather than having to parse through an immense data stream looking for the start of where each frame would be, the program logic only has to look at the circular event buffer in DRAM 714 and it can tell where the start of each frame is and the frame type. This approach saves a large amount of CPU power, keeping the real time requirements of the CPU 713 small. The CPU 713 does not have to

be very fast at any point in time. The Media Switch 701 gives the CPU 713 as much time as possible to complete tasks. The parsing mechanism 705 and event queue 708 decouple the CPU 713 from parsing the audio, video, and buffers and the real time nature of the streams, which allows for lower costs. It also allows the use of a bus structure in a CPU environment that operates at a much lower clock rate with much cheaper memory than would be required otherwise.

The CPU 713 has the ability to queue up one DMA transfer and can set up the next DMA transfer at its leisure. This gives the CPU 713 large time intervals within which it can service the DMA controller 709. The CPU 713 may respond to a DMA interrupt within a larger time window because of the large latency allowed. MPEG streams, whether extracted from an MPEG2 Transport or encoded from an analog TV signal, are typically encoded using a technique called Variable Bit Rate encoding (VBR). This technique varies the amount of data required to represent a sequence of images by the amount of movement between those images. This technique can greatly reduce the required bandwidth for a signal, however sequences with rapid movement (such as a basketball game) may be encoded with much greater bandwidth requirements. For example, the Hughes DirecTV satellite system encodes signals with anywhere from 1 to 10Mb/s of required bandwidth, varying from frame to frame. It would be difficult for any computer system to keep up with such rapidly varying data rates without this structure.

With respect to Fig. 8, the program logic within the CPU has three conceptual components: sources 801, transforms 802, and sinks 803. The sources 801 produce buffers of data. Transforms 802 process buffers of data and sinks 803 consume buffers of data. A transform is responsible for allocating and queuing the buffers of data on which it will operate. Buffers are allocated as if "empty" to sources of data, which give them back "full". The buffers are then queued and given to sinks as "full", and the sink will return the buffer "empty".

A source 801 accepts data from encoders, *e.g.*, a digital satellite receiver. It acquires buffers for this data from the downstream transform, packages the data into a buffer, then pushes the buffer down the pipeline as described above. The source object 801 does not know anything about the rest of the system. The sink 803 consumes buffers, taking a buffer from the upstream transform, sending the data to the decoder, and then releasing the buffer for reuse.

There are two types of transforms 802 used: spatial and temporal. Spatial transforms are transforms that perform, for example, an image convolution or compression/decompression on the buffered data that is passing through. Temporal transforms are used when there is no time relation that is expressible between buffers going in and buffers coming out of a system. Such a transform writes the buffer to a file 804 on the storage medium. The buffer is pulled out at a later time, sent down the pipeline, and properly sequenced within the stream.

Referring to Fig. 9, a C++ class hierarchy derivation of the program logic is shown. The TiVo Media Kernel (Tmk) 904, 908, 913 mediates with the operating system kernel. The kernel provides operations such as: memory allocation, synchronization, and threading. The TmkCore 904, 908, 913 structures memory taken from the media kernel as an object. It provides operators, new and delete, for constructing and deconstructing the object. Each object (source 901, transform 902, and sink 903) is multi-threaded by definition and can run in parallel.

The TmkPipeline class 905, 909, 914 is responsible for flow control through the system. The pipelines point to the next pipeline in the flow from source 901 to sink 903. To pause the pipeline, for example, an event called "pause" is sent to the first object in the pipeline. The event is relayed on to the next object and so on down the pipeline. This all happens asynchronously to the data going through the pipeline. Thus, similar to applications such as telephony, control of the flow of MPEG streams is asynchronous and separate from the streams themselves. This allows for a simple logic design that is at the same time powerful enough to support the features described previously, including pause, rewind, fast forward and others. In addition, this structure allows fast and efficient switching between stream sources, since buffered data can be simply discarded and decoders reset using a single event, after which data from the new stream will pass down the pipeline. Such a capability is needed, for example, when switching the channel being captured by the input section, or when switching between a live signal from the input section and a stored stream.

The source object 901 is a TmkSource 906 and the transform object 902 is a TmkXfrm 910. These are intermediate classes that define standard behaviors for the classes in the pipeline. Conceptually, they handshake buffers down the pipeline. The source object 901 takes data out of a physical data source, such as

the Media Switch, and places it into a PES buffer. To obtain the buffer, the source object 901 asks the down stream object in his pipeline for a buffer (allocEmptyBuf). The source object 901 is blocked until there is sufficient memory. This means that the pipeline is self-regulating; it has automatic flow control. When the source object 901 has filled up the buffer, it hands it back to the transform 902 through the pushFullBuf function.

The sink 903 is flow controlled as well. It calls nextFullBuf which tells the transform 902 that it is ready for the next filled buffer. This operation can block the sink 903 until a buffer is ready. When the sink 903 is finished with a buffer (*i.e.*, it has consumed the data in the buffer) it calls releaseEmptyBuf. ReleaseEmptyBuf gives the buffer back to the transform 902. The transform 902 can then hand that buffer, for example, back to the source object 901 to fill up again. In addition to the automatic flow-control benefit of this method, it also provides for limiting the amount of memory dedicated to buffers by allowing enforcement of a fixed allocation of buffers by a transform. This is an important feature in achieving a cost-effective limited DRAM environment.

The MediaSwitch class 909 calls the allocEmptyBuf method of the TmkClipCache 912 object and receives a PES buffer from it . It then goes out to the circular buffers in the Media Switch hardware and generates PES buffers. The MediaSwitch class 909 fills the buffer up and pushes it back to the TmkClipCache 912 object.

The TmkClipCache 912 maintains a cache file 918 on a storage medium. It also maintains two pointers into this cache: a push pointer 919 that shows where the next buffer coming from the source 901 is inserted; and a current pointer 920 which points to the current buffer used.

The buffer that is pointed to by the current pointer is handed to the Vela decoder class 916. The Vela decoder class 916 talks to the decoder 921 in the hardware. The decoder 921 produces a decoded TV signal that is subsequently encoded into an analog TV signal in NTSC, PAL or other analog format. When the Vela decoder class 916 is finished with the buffer it calls releaseEmptyBuf.

The structure of the classes makes the system easy to test and debug. Each level can be tested separately to make sure it performs in the appropriate manner, and the classes may be gradually aggregated to achieve the desired

functionality while retaining the ability to effectively test each object.

The control object 917 accepts commands from the user and sends events into the pipeline to control what the pipeline is doing. For example, if the user has a remote control and is watching TV, the user presses pause and the control object 917 sends an event to the sink 903, that tells it pause. The sink 903 stops asking for new buffers. The current pointer 920 stays where it is at. The sink 903 starts taking buffers out again when it receives another event that tells it to play. The system is in perfect synchronization; it starts from the frame that it stopped at.

The remote control may also have a fast forward key. When the fast forward key is pressed, the control object 917 sends an event to the transform 902, that tells it to move forward two seconds. The transform 902 finds that the two second time span requires it to move forward three buffers. It then issues a reset event to the downstream pipeline, so that any queued data or state that may be present in the hardware decoders is flushed. This is a critical step, since the structure of MPEG streams requires maintenance of state across multiple frames of data, and that state will be rendered invalid by repositioning the pointer. It then moves the current pointer 920 forward three buffers. The next time the sink 903 calls nextFullBuf it gets the new current buffer. The same method works for fast reverse in that the transform 902 moves the current pointer 920 backwards.

A system clock reference resides in the decoder. The system clock reference is sped up for fast play or slowed down for slow play. The sink simply asks for full buffers faster or slower, depending on the clock speed.

With respect to Fig. 10, two other objects derived from the TmkXfrm class are placed in the pipeline for disk access. One is called TmkClipReader 1003 and the other is called TmkClipWriter 1001. Buffers come into the TmkClipWriter 1001 and are pushed to a file on a storage medium 1004. TmkClipReader 1003 asks for buffers which are taken off of a file on a storage medium 1005. A TmkClipReader 1003 provides only the allocEmptyBuf and pushFullBuf methods, while a TmkClipWriter 1001 provides only the nextFullBuf and releaseEmptyBuf methods. A TmkClipReader 1003 therefore performs the same function as the input, or "push" side of a TmkClipCache 1002, while a TmkClipWriter 1001 therefore performs the same function as the output, or "pull" side of a TmkClipCache 1002.

Referring to Fig. 11, a preferred embodiment that accomplishes multiple functions is shown. A source 1101 has a TV signal input. The source sends data to a PushSwitch 1102 which is a transform derived from TmkXfrm. The PushSwitch 1102 has multiple outputs that can be switched by the control object 1114. This means that one part of the pipeline can be stopped and another can be started at the users whim. The user can switch to different storage devices. The PushSwitch 1102 could output to a TmkClipWriter 1106, which goes onto a storage device 1107 or write to the cache transform 1103.

An important feature of this apparatus is the ease with which it can selectively capture portions of an incoming signal under the control of program logic. Based on information such as the current time, or perhaps a specific time span, or perhaps via a remote control button press by the viewer, a TmkClipWriter 1106 may be switched on to record a portion of the signal, and switched off at some later time. This switching is typically caused by sending a "switch" event to the PushSwitch 1102 object.

An additional method for triggering selective capture is through information modulated into the VBI or placed into an MPEG private data channel. Data decoded from the VBI or private data channel is passed to the program logic. The program logic examines this data to determine if the data indicates that capture of the TV signal into which it was modulated should begin. Similarly, this information may also indicate when recording should end, or another data item may be modulated into the signal indicating when the capture should end. The starting and ending indicators may be explicitly modulated into the signal or other information that is placed into the signal in a standard fashion may be used to encode this information.

With respect to Fig. 12, an example is shown which demonstrates how the program logic scans the words contained within the closed caption (CC) fields to determine starting and ending times, using particular words or phrases to trigger the capture. A stream of NTSC or PAL fields 1201 is presented. CC bytes are extracted from each odd field 1202, and entered in a circular buffer 1203 for processing by the Word Parser 1204. The Word Parser 1204 collects characters until it encounters a word boundary, usually a space, period or other delineating character. Recall from above, that the MPEG audio and video segments are collected into a series of fixed-size PES buffers. A special segment is added to each PES buffer to hold the words extracted from the CC

field 1205. Thus, the CC information is preserved in time synchronization with the audio and video, and can be correctly presented to the viewer when the stream is displayed. This also allows the stored stream to be processed for C C information at the leisure of the program logic, which spreads out load, reducing cost and improving efficiency. In such a case, the words stored in the special segment are simply passed to the state table logic 1206.

During stream capture, each word is looked up in a table 1206 which indicates the action to take on recognizing that word. This action may simply change the state of the recognizer state machine 1207, or may cause the state machine 1207 to issue an action request, such as "start capture", "stop capture", "phrase seen", or other similar requests. Indeed, a recognized word or phrase may cause the pipeline to be switched; for example, to overlay a different audio track if undesirable language is used in the program.

Note that the parsing state table 1206 and recognizer state machine 1207 may be modified or changed at any time. For example, a different table and state machine may be provided for each input channel. Alternatively, these elements may be switched depending on the time of day, or because of other events.

Referring to Fig. 11, a PullSwitch is added 1104 which outputs to the sink 1105. The sink 1105 calls nextFullBuf and releaseEmptyBuf to get or return buffers from the PullSwitch 1104. The PullSwitch 1104 can have any number of inputs. One input could be an ActionClip 1113. The remote control can switch between input sources. The control object 1114 sends an event to the PullSwitch 1104, telling it to switch. It will switch from the current input source to whatever input source the control object selects.

An ActionClip class provides for sequencing a number of different stored signals in a predictable and controllable manner, possibly with the added control of viewer selection via a remote control. Thus, it appears as a derivative of a TmkXfrm object that accepts a "switch" event for switching to the next stored signal.

This allows the program logic or user to create custom sequences of video output. Any number of video segments can be lined up and combined as if the program logic or user were using a broadcast studio video mixer. TmkClipReaders 1108, 1109, 1110, are allocated and each is hooked into the

PullSwitch 1104. The PullSwitch 1104 switches between the TmkClipReaders 1108, 1109, 1110 to combine video and audio clips. Flow control is automatic because of the way the pipeline is constructed. The Push and Pull Switches are the same as video switches in a broadcast studio.

The derived class and resulting objects described here may be combined in an arbitrary way to create a number of different useful configurations for storing, retrieving, switching and viewing of TV streams. For example, if multiple input and output sections are available, one input is viewed while another is stored, and a picture-in-picture window generated by the second output is used to preview previously stored streams. Such configurations represent a unique and novel application of software transformations to achieve the functionality expected of expensive, sophisticated hardware solutions within a single cost-effective device.

With respect to Fig. 13, a high-level system view is shown which implements a VCR backup. The Output Module 1303 sends TV signals to the VCR 1307. This allows the user to record TV programs directly on to video tape. The invention allows the user to queue up programs from disk to be recorded on to video tape and to schedule the time that the programs are sent to the VCR 1307. Title pages (EPG data) can be sent to the VCR 1307 before a program is sent. Longer programs can be scaled to fit onto smaller video tapes by speeding up the play speed or dropping frames.

The VCR 1307 output can also be routed back into the Input Module 1301. In this configuration the VCR acts as a backup system for the Media Switch 1302. Any overflow storage or lower priority programming is sent to the VCR 1307 for later retrieval.

The Input Module 1301 can decode and pass to the remainder of the system information encoded on the Vertical Blanking Interval (VBI). The Output Module 1303 can encode into the output VBI data provided by the remainder of the system. The program logic may arrange to encode identifying information of various kinds into the output signal, which will be recorded onto tape using the VCR 1307. Playing this tape back into the input allows the program logic to read back this identifying information, such that the TV signal recorded on the tape is properly handled. For example, a particular program may be recorded to tape along with information about when it was recorded, the source network, etc.

When this program is played back into the Input Module, this information can be used to control storage of the signal, presentation to the viewer, etc.

One skilled in the art will readily appreciate that such a mechanism may be used to introduce various data items to the program logic which are not properly conceived of as television signals. For instance, software updates or other data may be passed to the system. The program logic receiving this data from the television stream may impose controls on how the data is handled, such as requiring certain authentication sequences and/or decrypting the embedded information according to some previously acquired key. Such a method works for normal broadcast signals as well, leading to an efficient means of providing non-TV control information and data to the program logic.

Additionally, one skilled in the art will readily appreciate that although a VCR is specifically mentioned above, any multimedia recording device (*e.g.*, a Digital Video Disk-Random Access Memory (DVD-RAM) recorder) is easily substituted in its place.

Although the invention is described herein with reference to the preferred embodiment, one skilled in the art will readily appreciate that other applications may be substituted for those set forth herein without departing from the spirit and scope of the present invention. For example, the invention can be used in the detection of gambling casino crime. The input section of the invention is connected to the casino's video surveillance system. Recorded video is cached and simultaneously output to external VCRs. The user can switch to any video feed and examine (*i.e.*, rewind, play, slow play, fast forward, etc.) a specific segment of the recorded video while the external VCRs are being loaded with the real-time input video.

Video Stream Tag Architecture

Referring again to Fig. 12, tags are abstract events which occur in a television stream 1201. They may be embedded in the VBI of an analog signal, or in a private data channel in an MPEG2 multiplex. As described above, tags can be embedded in the closed caption (CC) fields and extracted into a circular buffer 1203 or memory allocation schema. The word parser 1204 identifies unique tags during its scan of the CC data. Tags are interspersed with the standard CC control codes. Tags may also be generated implicitly, for instance, based on the

current time and program being viewed.

The invention provides a mechanism called the TiVo Video Tag Authoring (TVTAG) system for inserting tags (TiVo tags) into a video stream prior to broadcast. With respect to Figs. 14, 16, and 17, the TVTAG system consists of a video output source 1401, a compatible device for inserting Vertical Blanking Interval (VBI) closed-captioning information and outputting captioned video 1402, a video monitor 1405, and a software program for controlling the VBI insertion device to incorporate tag data objects in the form of closed-caption information in the video stream 1406. The tagged video is retransmitted immediately 1404 or stored on a suitable medium 1403 for later transmission.

The TVTAG software 1406, in its most basic implementation, is responsible for controlling the VBI Insertion device 1402. The TVTAG software 1406 communicates with the VBI insertion device 1402 by means of standard computer interfaces and device control code protocols. When an operator observing the video monitor 1405 determines that the desired tag insertion point has been reached, he presses a key, causing the TiVo tag data object to be generated, transmitted to the VBI insertion device 1402, and incorporated in the video stream for transmission1404 or storage 1403.

The TVTAG software has the additional capability of controlling the video input source 1401 and the video output storage device 1403. The operator selects the particular video 1602 and has the ability to pause the video input stream to facilitate overlaying a graphic element 1702 on the monitor, and positioning it by means of a pointing device, such as a mouse. The positioning of the graphic element 1702 is also accomplished through the operator interface 1601. The operator inputs the position of the graphic using the X position 1605 and the Y position 1604.

The graphic element and positioning information are then incorporated in the TiVo tag data object (discussed below) and the time-code or frame of the video noted. When the operator is satisfied, playback and record are resumed. The tag is then issued through the insertion device with the highest degree of accuracy.

Referring to Fig. 15, in another preferred embodiment of the TVTAG system, the software program takes the form of a standard Internet protocol Web page

displayed to operator(s) 1505. The Web page causes the TiVo tag object to be generated by a script running on a remote server 1504. The server 1504 controls the VBI insertion device 1502, the video source 1501, and recording devices 1503. The remote operator(s) 1505 can receive from the server 1504 a low or high-bandwidth version of the video stream for use as a reference for tag insertion. Once the necessary tag data object information has been generated and transmitted, it can be batch-processed at a later time by the server 1504.

Another preferred embodiment of the invention integrates the software with popular non-linear video editing systems as a "plug-in", thereby allowing the TiVo tag data objects to be inserted during the video production process. In this embodiment, the non-linear editing system serves as the source and storage system controller and also provides graphic placement facilities, allowing frame-accurate placement of the TiVo tag data object.

With respect to Fig. 18, tags are integrated into the video stream before or at the video source 1801. The video stream is then transmitted via satellite 1802, cable or other terrestrial transmission method 1803. The receiver 1804 receives the video stream, recognizes the tags and performs the appropriate actions in response to the tags. The viewer sees the resultant video stream via the monitor or television set 1805.

The invention provides an architecture that supports taking various actions based on tags in the video stream. Some examples of the flexibility that TiVo tags offer are:

- It is desirable to know when a network promotion is being viewed so that the viewer might be presented with an option to record the program at some future time. TiVo tags are added into the promotion that indicate the date, time, and channel when the program airs. Active promos are described in further detail below.

- A common problem is the baseball game overrun problem. VCRs and Digital Video Recorders (DVR) cut off the end of the baseball game whenever the game runs over the advertised time slot. A TiVo tag is sent in the video stream indicating that the recording needs to continue. A TiVo tag is also sent telling the system to stop the recording because the game has ended.

- Boxing matches often end abruptly, causing VCRs and DVRs to record fill-in programs for the rest of the reserved time period. A TiVo tag is sent to indicate that the program has ended, telling the system to stop the recording.

- Referring to Fig. 19, advertisements are tagged so a locally or remotely stored advertisement might be shown instead of a national or out of the area advertisement. Within the video stream 1901, the program segment 1902 (commercial or other program segment) to be overlaid is tagged using techniques such as the TVTAG system described above. The TiVo tags tell the invention 1905 the start and end points of the old program segment 1902. A single tag 1903 can be added that tells the invention 1905 the duration of the old program segment 1902 or a tag is added at the beginning 1903 and end 1904 of the old program segment to indicate the start and end of the segment 1902. When the TiVo tag is detected, the invention 1905 finds the new program segment 1906 and simply plays it back in place of the old program segment 1902, reverting to the original program 1901 when playback is completed. The viewer 1907 never notices the transition.

There are three options at this point:

1) The system 1905 can continue to cache the original program, so if the viewer 1907 rewinds the program 1901 and plays it again, he sees the overlaid segment;

2) The old program segment 1902 is replaced in the cache too, so the viewer never sees the overlaid segment; or

3) The system caches the original segment 1902 and reinterprets the tags on playback. However, without intelligent tag prefetching, this only works correctly if the viewer backs up far enough so the system sees the first tag in the overlaid segment.

This problem is solved by adding the length of the old program segment to the start 1903 and end 1904 tag. Another approach is to match tags so that the start tag 1903 identifies the end tag 1904 to the system. The system 1905 knows that it should be looking for another tag when it fast forwards or rewinds over one of the tags. The pair of tags 1903, 1904 include a unique identifier. The system 1905 can then search ahead or

behind for the matching tag and replace the old program. There is a limit to the amount of time or length of frames that the system can conduct the prefetch. This can be included in the tag or standardized. Including the limit in the tag is the most flexible approach.

The program segment to be played back is selected based, for example, on locale, the time of day, program material, or on the preference engine (described in Application No. 09/422,121 owned by the Applicant). Using the preference engine, the appropriate program segment from local or server storage 1906 is selected according to the viewer's profile. The profile contains the viewer's viewing habits, program preferences, and other personal information. The stored program segments 1906 have program objects describing their features as well, which are searched for best match versus the preference vector.

Clearly, there must be a rotation mechanism among commercials to avoid ad burnout. The preference vector can be further biased by generating an error vector versus the program data for the currently viewed program, and using this error vector to bias the match against the commercial inventory on disk 1906. For example, if the viewer is watching a soap opera and the viewer's preference vector is oriented towards sports shows, then the invention will select the beer commercial in favor of the diaper commercial.

A tag can also be used to make conditional choices. The tag contains a preference weighting of its own. In this case, the preference weighting is compared to the preference vector and a high correlation causes the invention to leave the commercial alone. A low correlation invokes the method above.

NOTE: In all of these cases the system 1905 has more than enough time to make a decision. The structure of the pipeline routinely buffers 1/2 second of video, giving lots of time between input and output to change the stream. If more time is needed, add buffering to the pipeline. If playing back off disk, then the system creates the same time delay by reading ahead in the stream.

Also note that commercials can also be detected using the method described in Application No. 09/187,967 entitled "Analog Video Tagging and Encoding System," also owned by the Applicant. The same type of substitution described above can be used when tags described the aforementioned

application are used.

- With respect to Figs. 19 and 22, tags allow the incorporation of commercial "zapping." Since tags can be used to mark the beginning 1903 and ending 1904 points of a commercial, they can be skipped as well as preempted. The viewer simply presses the jump button 2205 on the remote control 2201. The system searches for the end tag and resumes playback at the frame following the frame associated with the tag. The number of commercials skipped is dependent upon the amount of video stream buffered.

  Depending on the viewer's preset preferences, the system 1905 itself can skip commercials on live or prerecorded programs stored in memory 1906. Skipping commercials on live video just requires a larger amount of buffering in the pipeline as described above. Allowing the system to skip commercials on recorded programs presents the viewer with a continuous showing of the program without any commercial interruptions.

- Tags are added to program material to act as indexes. The viewer, for example, can jump to each index within the program by pressing the jump button 2205 on the remote control 2201.

- Tags are also used for system functions. As noted above, the system locally stores program material for its own use. The system 1905 must somehow receive the program material. This is done by tuning in to a particular channel at off hours. The system 1905 searches for the tag in the stream 1901 that tells it to start recording. The recording is comprised of a number of program segments delimited by tags 1903, 1904 that identify the content and possibly a preference vector. A tag at the end of the stream tells the system 1905 to stop recording. The program segments are stored locally 1906 and indexed for later use as described above.

The invention incorporates the following design points:

- The design provides for a clear separation of mechanism and policy.

- Internally, tags are viewed as abstract events which trigger policy modules. Mapping of received tag information to these internal abstractions is the

responsibility of the source pipeline object.

- Abstract tags are stored in the PesBuf stream as if they were just another segment. This allows the handling of arbitrary sized tags with precise timing information. It also allows tags to persist as part of recorded programs, so that proper actions are taken no matter when the program is viewed.

- Tags may update information about the current program, future programs, etc. This information is preserved for recorded programs.

- Tags can be logged as they pass through the system. It also possible to upload this information. It may not be necessary to preserve all information associated with a tag.

- Tags can be generated based on separate timelines. For example, using a network station log to generate tags based on time and network being viewed. Time-based tags are preserved in recorded streams.

Time-Based Tags

Referring to Fig. 20, time-based tags are handled by a Time-based Tag Recognizer 2012. This object 2012 listens for channel change events and, when a known network is switched to, attempts to retrieve a "time log" for that network. If one is present, the object 2012 builds a tag schedule based on the current time. As the time occurs for each tag, the object 2012 sends an event to the source object 2001 indicating the tag to be inserted. The source object 2001 inserts the tag into the next available position in the current PesBuf under construction. The next "available" position may be determined based on frame boundaries or other conditions.

The Role of the Source Object

The source object 2001 is responsible for inserting tags into the PesBuf stream it produces. This is assuming there are separate source objects for analog input and digital TV sources.

There are a number of different ways that tags may appear in an analog stream:

- Within the EDS field.
- Implicitly using the CC field.
- Modulated onto the VBI, perhaps using the ATVEF specification.
- Time Based

In a digital TV stream, or after conversion to MPEG from analog:

- In-band, using TiVo Tagging Technology.
- MPEG2 Private data channel.
- MPEG2 stream features (frame boundaries, etc.).
- Time-based tags.

The source object 2001 is not responsible for parsing the tags and taking any actions. Instead, the source object 2001 should solely be responsible for recognizing potential tags in the stream and adding them to the PesBuf stream.

Tag Recognition and Action

Conceptually, all tags may be broken up into two broad groups: those that require action upon reception, such as recording a program; and those that require action upon presentation, *i.e.*, when the program is viewed.

Reception Tag Handling

Tags that require action upon reception are handled as follows: a new Reception Tag Mechanism subclass 2003 of the TmkPushSwitch class 2002 is created. As input streams pass through this class 2003 between the source object 2001 and the program cache transform 2013, the class 2003 recognizes reception tags and takes appropriate actions.

Reception tags are generally handled once and then disabled.

Presentation Tag Handling

Tags that require actions upon presentation are handled as follows: a new Presentation Tag Mechanism subclass 2007 of the TmkPullSwitch class 2008 is created. As output streams pass through this class 2007 between the program cache transform 2013 and the sink object 2011, the class 2007 recognizes

presentation tags and takes appropriate actions.

Tag Policy Handling

Tag reception handling is only permitted if there is a TagReceptionPolicy object 2009 present for the current channel. Tag presentation handling is only permitted if there is a TagPresentationPolicy object 2010 for the source channel.

The TagPolicy objects describe which tags are to be recognized, and what actions are allowed.

When an input channel change occurs, the reception tag object is notified, and it fetches the TagReceptionPolicy object 2009 (if any) for that channel, and obeys the defined policy.

When an output channel change occurs, the presentation tag object is notified, and it fetches the TagPresentationPolicy object 2010 (if any) for that channel, and obeys the defined policy.

Tag Logging

The reception of tags may be logged into the database. This only occurs if a TagReceptionPolicy object 2009 is present, and the tag logging attribute is set. As an example, the logging attribute might be set, but no reception actions allowed to be performed. This allows passive logging of activity in the input stream.

Pipeline Processing Changes

It is important to support updates of information about the current showing. The following strategy is proposed:

- Whenever the input source is changed or a new showing starts, a copy is made of the showing object, and all further operations in the pipeline work off this copy.

- Update tags are reception tags; if permitted by policy, the copied showing object is updated.

- If the current showing is to be recorded, the copy of the showing object is saved with it, so that the saved program has the proper information saved with it.

- The original showing object is not modified by this process.

- The recorder must be cognizant of changes to the showing object, so that it doesn't, for instance, cut off the baseball game early.

Tag Interpretation vs. Tag State Machine

Tags are extremely flexible in that, once the TagPolicy object has been used to identify a valid tag, standardized abstract tags are interpreted by the Tag Interpreter 2005 and operational tags are executed by the TiVo Tag State Machine 2006. Interpreted tags trigger a predefined set of actions. Each set of actions have been preprogrammed into the system.

State machine tags are operational tags that do not carry executable code, but perform program steps. This allows the tag originator to combine these tags to perform customized actions on the TiVo system. State machine tags can be used to achieve the same results as an interpreted tag, but have the flexibility to dynamically change the set of actions performed.

Abstract Interpreted Tags

The set of available abstract tags is defined in a table called the Tag/Action table. This table is typically stored in a database object. There are a small number of abstract actions defined. These actions fall into three general categories:

- Viewer visible actions (may include interaction).
- Meta-information about the stream (channel, time, duration, etc.).
- TiVo control tags.

Tags which cause a change to the on-disk database, or cause implicit recording, must be validated. This is accomplished through control tags.

Viewer Visible Tags

- Menu

This tag indicates that the viewer is to be presented with a choice. The data associated with the tag indicates what the choice is, and other interesting data, such as presentation style. A menu has an associated inactivity timeout.

The idea of the menu tag is that the viewer is offered a choice. If the viewer isn't present, or is uninterested, the menu should disappear quickly. The menu policy may or may not be to pause the current program. The presentation of the menu does not have to be a list.

- Push Alternate Program Conditional

This tag indicates that some alternate program should be played if some condition is true. The condition is analyzed by the policy module. It may always be true.

- Pop Alternate Program Conditional

This tag reverts to the previous program. If a program ends, then the alternate program stack is popped automatically. All alternate programs are popped if the channel is changed or the viewer enters the TiVo Central menu area.

Alternate programs are a way of inserting arbitrary sequences into the viewed stream. The conditional data is not evaluated at the top level. Instead, the policy module must examine this data to make choices. This, for example, can be used to create "telescoping" ads.

- Show Indicator Conditional

This tag causes an indication to be drawn on the screen. Indicators are named, and the set of active indicators may be queried at any time. The tag or tag policy may indicate a timeout value at which time the indicator is derived.

- Clear Indicator Conditional

This tag causes an active indication to be removed. All indicators are cleared if the channel is changed or the viewer enters the TiVo Central menu area.

Indicators are another way to offer a choice to the viewer without interrupting program flow. They may also be used to indicate conditions in the stream that may be of interest. For example, "Active Promo" is created by providing a program object ID as part of the tag data, allowing that program to be selected. If the viewer hits a particular key while the indicator is up, then the program is scheduled for recording.

Meta-Information Tags

- Current Showing Information

This tag is a general bucket for information about the current showing. Each tag typically communicates one piece of information, such as the start time, end time, duration, etc. This tag can be used to "lengthen" a recording of an event.

- Future Showing Information

This tag is similar to the above, but contains information about a future showing. There are two circumstances of interest:

- The information refers to some showing already resident in the database. The database object is updated as appropriate.
- The information refers to a non-existent showing. A new showing object is created and initialized from the tag.

TiVo Control Tags

- Authorize Modification

This tag is generally encrypted with the current month's security key. The lifetime of the authorization is set by policy, probably to an hour or two. Thus, the tag needs to be continually rebroadcast if modifications to local TiVo system states are permitted.

The idea of this tag is to avoid malicious (or accidental) attacks using inherently insecure tag mechanisms such as EDS. If a network provides EDS information,

we first want to ensure that their tags are accurate and that attacks on the tag delivery system are unlikely. Then, we would work with that network to provide an authorization system that carouseled authorization tags on just that network. Unauthorized tags should never be inserted into the PES stream by the source object.

- Record Current Conditional

This tag causes the current program to be saved to disk starting from this point. The recording will cease when the current program ends.

- Stop Recording Current Conditional

This tag ceases recording of the current program.

- Record Future Conditional

A showing object ID is provided (perhaps just sent down in a Future Showing tag). The program is scheduled for recording at a background priority lower than explicit viewer selections.

- Cancel Record Future Conditional

A showing object ID is provided. If a recording was scheduled by a previous tag for that object, then the recording is canceled.

These tags, and the Future Showing tag, may be inserted in an encrypted, secure format. The source object will only insert these tags in the PES stream if they are properly validated.

One of the purposes of these tags is to automatically trigger recording of TiVo inventory, such as loopsets, advertisements, interstitials, etc. A later download would cause this inventory to be "installed" and available.

- Save File Conditional

This tag is used to pass data through the stream to be stored to disk. For instance, broadcast Web pages would be passed through

this mechanism.

- Save Object Conditional

This tag is used to pass an object through the stream to be stored to disk. Storing the object follows standard object updating rules.

The following is an example of an implementation using presentation tags inserted into the Closed Captioning (CC) part of a stream. The CC part of the stream was chosen because it is preserved when a signal is transmitted and digitized and decoded before it reaches the user's receiver. There are no guarantees on the rest of the VBI signal. Many of the satellite systems strip out everything except the closed captioning when encoding into MPEG-2.

There is a severe bandwidth limitation on the CC stream. The data rate for the CC stream is two 7-bit bytes every video frame. Furthermore, to avoid collision with the control codes, the data must start at 0x20, thus effectively limiting it to about 6.5-bit bytes (truncate to 6-bit bytes for simplicity). Therefore, the bandwidth is roughly 360 bits/second. This rate gets further reduced if the channel is shared with real CC data. In addition, extra control codes need to be sent down to prevent CC-enabled televisions from attempting to display the TiVo tags as CC text.

Basic Tag Layout

This section describes how the tags are laid out in the closed captioning stream. It assumes a general familiarity with the closed captioning specification, though this is not crucial.

Making Tags Invisible

A TiVo Tag placed in a stream should not affect the display on a closed captioning enabled television. This is achieved by first sending down a "resume caption loading" command (twice for fault tolerance), followed by a string of characters that describes the tag followed by an "erase nondisplayed memory" command (twice for fault tolerance). What this does is to load text into offscreen memory, and then clear the memory. A regular TV with closed captioning enabled will not display this text (as per EIA-701 standard).

This works as long as the closed captioning decoder is not in "roll-up" or "scrolling" mode. In this mode, a "resume caption loading" command would cause the text to be erased. To solve this problem, TiVo Tags will be accepted and recognized even if they are sent to the second closed captioning channel. This way, even if closed captioning channel 1 is set up with scrolling text, we can still send the tag through closed captioning channel 2.

Tag Encoding

The text sent with a TiVo Tag consists of the letters "Tt", followed by a single character indicating the length of the tag, followed by the tag contents, followed by a CRC for the tag contents. The letters "Tt" are sufficiently unique that it is unlikely to encounter these in normal CC data. Furthermore, normal CC data always starts with a position control code to indicate where on the screen the text is displayed. Since we are not displaying onscreen, there is no need for this positioning data. Therefore, the likelihood of encountering a "Tt" immediately after a "resume caption loading" control code is sufficiently rare that we can almost guarantee that this combination is a TiVo tag (though the implementation still will not count on this to be true).

The single character indicating the length of the tag is computed by adding the tag length to 0x20. If the length is 3 characters, for example, then the length character used is 0x23 ('#'). So as not to limit the implementation to a length of 95 (since there are only 96 characters in the character set), the maximum length is defined as 63. If longer tags are needed, then an interpretation for the other 32 possible values for the length character can be added.

The possible values for the tag itself are defined in the Tag Types section below.

The CRC is the 16 bit CRC-CCITT (i.e., polynomial = x^16 + x^12 + x^5 + 1). It is placed in the stream as three separate characters. The first character is computed by adding 0x20 to the most significant six bits of the CRC. The next character is computed by adding 0x20 to the next six bits of the CRC. The last character is computed by adding 0x20 to the last four bits of the CRC.

Tag Types

This section details an example of a TiVo Tag. Note that every tag sequence begins with at least one byte indicating the tag type.

iPreview Tag

With respect to Fig. 17, an iPreview tag contains four pieces of information. The first is the 32 bit program ID of the program being previewed. The second contains how much longer the promotion is going to last. The third piece is where on the screen 1701 to place an iPreview alert 1702 and the last piece is what size iPreview alert to use.

The screen location for the iPreview alert is a fraction of the screen resolution in width and height. The X coordinate uses 9 bits to divide the width; so the final coordinate is given as: `X = (x_resolution/511) * xval`. If the xval is given as 10, on a 720 x 486 screen (using CCIR656 resolution), the X coordinate would be 14. The Y coordinate uses 8 bits to divide the height, so the final coordinate is given as: `Y = (y_resolution/255) * yval`. The X,Y coordinates indicate the location of the upper-left corner of the bug graphic.

If the value of X and Y are set to the maximum possible values (i.e., x=511, y=255), then this indicates that the author is giving the system the job of determining its position. The system will place the bug at a predetermined default position. The rationale for using the max values to indicate the default position is that it is never expected that a "real" position will be set to these values since that would put the entire bug graphic offscreen.

The size field is a four bit number that indicates what size any alert graphic should be. The 16 possible values of this field correspond to predefined graphic sizes that the settop boxes should be prepared to provide.

The timeout is a ten bit number indicating the number of frames left in the promotion. This puts a 34 second lifetime limit on this tag. If a promotion is longer, then the tag needs to be repeated. Note that the timeout was "artificially limited" to 10 bits to limit exposure to errors. This is to limit the effect it will have on subsequent commercials if an author puts a malformed timeout in the tag.

The version is a versioning number used to identify the promo itself. Instead of

bit-packing this number (and thus limiting it to 6 bits), the full closed captioning character set is used, which results in 96 possibilities instead of 64 (2^6). The version number thus needs to be within the range 0-95.

The reserved character is currently unused. This character needs to exist so that the control codes end up properly aligned on the 2-byte boundaries.

The first character of an iPreview tag is always "i".

All of the data fields are packed together on a bit boundary, and then broken into six bit values which are converted into characters (by adding 0x20) and transmitted. The order of the fields are as follows:

- 32 bits: program ID
- 9 bits: X location
- 8 bits: Y location
- 4 bits: graphic size
- 10 bits: timeout
- 1 character: version
- 1 character: reserved

The data fields total 66 bits which requires 11 characters to send + 1 character for version and 1 character for reserved. The exact contents of each character are:

1) 0x20 + ID[31:26]
2) 0x20 + ID[25:20]
3) 0x20 + ID[19:14]
4) 0x20 + ID[13:8]
5) 0x20 + ID[7:2]
6) 0x20 + ID[1:0] X[8:5]
7) 0x20 + X[4:0] Y[7]
8) 0x20 + Y[6:1]
9) 0x20 + Y[0] size[3:0]
10) 0x20 + Y[0] size[3:0] timeout[9]
11) 0x20 + timeout[8:3]
12) 0x20 + timeout[2:0]
13) 0x20 + version
14) reserved

Including the first character "i", the length of the iPreview tag is 14 characters + 3 CRC characters. With the tag header (3 characters), this makes a total length of 20 characters which can be sent down over 10 frames. Adding another 4 frames for sending "resume caption loading" twice and "erase nondisplayed memory" twice means an iPreview tag will take 14 frames (0.47 seconds) to broadcast.

A complete iPreview tag consists of:

Resume caption loading Resume caption loading T t 1 (0x20 + 17 = 0x31 = 0110001= "1") i <13 character iPreview tag> 3 character CRC Erase nondisplayed memory Erase nondisplayed memory

Parity debugging character

Currently, the parity bit is being used as a parity bit. However, since a CRC is already included, there is no need for the error-checking capabilities of the parity bit. Taking this a step further, the parity bit can be used in a clever way. Since a closed captioning receiver should ignore any characters with an incorrect parity bit, a better use of the limited bandwidth CC channel can be had by intentionally using the wrong parity. This allows the elimination of the resume caption loading and erase nondisplayed memory characters, as well as making it easier to "intersperse" TiVo tags among existing CC data.

iPreview Viewer Interaction

Referring to Figs. 17, 20, 21 and 22, the iPreview tag causes the Tag Interpreter 2005 to display the iPreview alert 1702 on the screen 1701. The iPreview alert 1702 tells the viewer that an active promo is available and the viewer can tell the TiVo system to record the future showing. The viewer reacts to the iPreview alert 1702 by pressing the select button 2204 on the remote control 2201.

The Tag Interpreter 2005 waits for the user input. Depending on the viewer's preset preferences, the press of the select button 2204 results in the program automatically scheduled by the Tag Interpreter 2005 for recording, resulting in a one-touch record, or the viewer is presented with a record options screen 2101. The viewer highlights the record menu item 2102 and presses the select button 2204 to have the program scheduled for recording.

The tag itself has been interpreted by the Tag Interpreter 2005. The Tag Interpreter 2005 waits for any viewer input through the remote control 2201. Once the viewer presses the select button 2204, the Tag Interpreter 2005 tells the TiVo system to schedule a recording of the program described by the 32 bit program ID in the iPreview tag.

With respect to Figs. 20, 22, and 23, the iPreview tag is also used for other purposes. Each use is dictated by the context of the program material and the screen icon displayed. Obviously the system cannot interpret the program material, but the icon combined with the program ID tell the Tag Interpreter 2005 what action to take. Two examples are the generation of a lead and a sale.

The process of generating a lead occurs when, for example, a car ad is being played. An iPreview icon appears 2301 on the screen and the viewer knows that he can press the select button 2204 to enter an interactive menu.

A menu screen 2302 is displayed by the Tag Interpreter 2005 giving the user the choice to get more information 2303 or see a video of the car 2304. The viewer can always exit by pressing the live TV button 2202. If the viewer selects get more information 2303 with the up and down arrow button 2203 select button 2204, then the viewer's information is sent to the manufacturer 2305 by the Tag Interpreter 2005, thereby generating a lead. The viewer returns to the program by pressing the select button 2204.

Generating a sale occurs when a product, *e.g.*, a music album ad, is advertised. The iPreview icon 2301 appears on the screen. The viewer presses the select button 2204 and a menu screen 2307 is displayed by the Tag Interpreter 2005.

The menu screen 2307 gives the viewer the choice to buy the product 2308 or to exit 2309. If the viewer selects yes 2308 to buy the product, then the Tag Interpreter 2005 sends the order to the manufacturer with the viewer's purchase information 2310. If this were a music album ad, the viewer may also be presented with a selection to view a music video by the artist.

Whenever the system returns the viewer back to the program, it returns to the exact point that the viewer had originally exited from. This gives the viewer a sense of continuity.

The concept of redirection is easily expanded to the Internet. The iPreview icon will appear as described above. When the viewer presses the select button 2204 on the remote control 2201, a Web page is then displayed to the viewer. The viewer then interacts with the Web page and when done, the system returns the viewer back to the program that he was watching at the exact point from which the viewer had exited.

Using the preference engine as noted above, the information shown to the viewer during a lead or sale generation is easily geared toward the specific viewer. The viewer's viewing habits, program preferences, and personal information are used to select the menus, choices, and screens presented to the viewer. Each menu, choice and screen has an associated program object that is compared to the viewer's preference vector.

For example, if a viewer is male and the promo is for Chevrolet, then when the viewer presses the select button, a still of a truck is displayed. If the viewer were female, then a still of a convertible would be displayed.

Note that the Tag State Machine 2006 described below is fully capable of performing the same steps as the Tag Interpreter 2005 in the above examples.

The TiVo Tag State Machine

Referring again to Fig. 20, a preferred embodiment of the invention provides a Tag State Machine (TSM) 2006 which is a mechanism for processing abstract TiVo tags that may result in viewer-visible actions by the TiVo Receiver.

A simple example is the creation of an active promo. As demonstrated above, an active promo is where a promotion for an upcoming show is broadcast and the viewer is immediately given the option of having the TiVo system record that program when it actually is broadcast.

Hidden complexities underlie this simple example: some indicator must be generated to alert the viewer to the opportunity; the indicator must be brought into view or removed with precision; accurate identification of the program in question must be provided; and the program within which the active promo appears may be viewed at a very different time then when it was broadcast.

Creation and management of the TiVo tags is also challenging. It is important to cause as little change as possible to existing broadcast practices and techniques. This means keeping the mechanism as simple as possible for both ease of integration into the broadcast stream and for robust and reliable operation.

Principles of Tags

As previously noted, it is assumed that the bandwidth available for sending tags is constrained. For example, the VBI has limited space available which is under heavy competition. Even in digital television signals, the amount of out-of-band data sent will be small since most consumers of the signal will be mainly focused on television programming options.

A tag is then a simple object of only a few bytes in size. More complex actions are built by sending multiple tags in sequence.

The nature of broadcast delivery implies that tags will get lost due to signal problems, sunspots, etc. The TSM incorporates a mechanism for handling lost tags, and insuring that no unexpected actions are taken due to lost tags.

In general, viewer-visible tag actions are relevant only to the channel on which they are received; it is assumed that tag state is discarded after a channel change.

Physical tags are translated into abstract tags by the source object 1901 receiving the physical tag. Tags are not "active agents" in that they carry no executable code; functioning the TSM may result in viewer-visible artifacts and changes, but the basic operation of the TiVo receiver system will remain unaffected by the sequence of tags. If tags could contain executable code, such as the Java byte streams contemplated by the ATVEF, the integrity of the TiVo viewing experience might be compromised by poorly written or malicious software.

All tag actions are governed by a matching policy object matched to the current channel. Any or all actions may be enabled or disabled by this object; the absence of a policy object suppresses all tag actions.

The Basic Abstract Tag

All abstract tags have a common infrastructure. The following components are present in any abstract tag:

- Tag Type (1 byte)

The type 0 is disallowed. The type 255 indicates an "extension" tag, should more than 254 tag values be required at some future time.

- Tag Sequence (1 byte)

This unsigned field is incremented for each tag that is part of a sequence. Tags which are not part of a sequence must have this field set to zero. A tag sequence of one indicates the start of a new sequence; a sequence may be any length conceptually, but it will be composed of segments of no more than 255 tags in order.

Each tag type has an implicit sequence length (which may be zero); the sequence number is introduced to handle dropouts or other forms of tag loss in the stream. In general, if a sequence error occurs, the entire tag sequence is discarded and the state machine reset.

Tags should be checksummed in the physical domain. If the checksum doesn't match, the tag is discarded by the source object. This will result in a sequence error and reset of the state machine.

- Tag Timestamp (8 bytes)

This is the synchronous time within the TV stream at which the tag was recognized. This time is synchronous to all other presentation times generated by the TiVo Receiver. This component is never sent, but is generated by the receiver itself.

- Tag Data Length (2 bytes)

This is the length of any data associated with the tag. The interpretation of this data is based on the tag type. The physical domain translator should perform some minimal error checking on the data.

The Tag State Machine

The TSM is part of the Tag Presentation Mechanism, which is in-line with video playback.

Conceptually, the TSM manages an abstract stack of integer values with at least 32 bits of precision, or sufficient size to hold an object ID. The object ID is abstract, and may or may not indicate a real object on the TiVo Receiver - it may otherwise need to be mapped to the correct object. The stack is limited in size to 255 entries to limit denial-of-service attacks.

The TSM also manages a pool of variables. Variables are named with a 2-byte integer. The variable name 0 is reserved. "User" variables may be manipulated by tag sequences; such variables lie between 1 and 2^15-1. "System" variables are maintained by the TSM, and contain values about the current TiVo Receiver, such as: the current program object ID; the TSM revision; and other useful information. These variables have names between 2^15 and 2^16-1. The number of user variables may be limited within a TSM; a TSM variable indicates what this limit is.

The tag data is a sequence of TSM commands. Execution of these commands begins when the tag is recognized and allowed. TSM commands are byte oriented and certain commands may have additional bytes to support their function.

The available TSM commands may be broken down into several classes:

Data Movement Commands

push_byte - push the byte following the command onto the stack.
push_short - push the short following the command onto the stack.
push_word - push the word following the command onto the stack.

Variable Access Commands

push_var - push the variable named in the 16-bit quantity following the command.
pop_var - pop into the variable named jn the 16-bit quantity following the

command.

copy_var - copy into the variable named in the 16-bit quantity following the command from the stack.

Stack Manipulation Commands

swap - swap the top two stack values.

pop - toss the top stack value.

Arithmetic Commands

add_byte - add the signed byte following the command to the top of stack.

add_short - add the signed short following the command to the top of stack.

add_word - add the signed word following the command to the top of stack.

and - and the top and next stack entries together, pop the stack and push the new value.

or - or the top and next stack entries together, pop the stack and push the new value.

Conditional Commands

(Unsigned comparisons only)

brif_zero - branch to the signed 16-bit offset following the command if the top of stack is zero.

brif_nz - branch to the signed 16-bit offset following the command if the top of stack is not zero.

brif_gt - branch to the signed 16-bit offset following the command if the top of stack is greater than the next stack entry.

brif_ge - branch to the signed 16-bit offset following the command if the top of stack is greater than or equal to the next stack entry.

brif_le - branch to the signed 16-bit offset following the command if the top of stack is less than or equal to the next stack entry.

brif_lt - branch to the signed 16-bit offset following the command if the top of stack is less than the next stack entry.

brif_set - branch to the signed 16-bit offset following the command if there are bits set when the top and next stack entries are ANDed together.

Action Commands

exec - execute tag action on the object ID named on top of stack.
fin - terminate tag taking no action.

System Variables

32768 (TAG) - value of current tag.

Times in GMT:

32769 (YEAR) - current year (since 0).
32770 (MONTH) - current month (1-12).
32771 (DAY) - day of month (1-31).
32772 (WDAY) - day of week (1-7, starts Sunday).
32773 (HOUR) - hour of the day (0-23).
32774 (MIN) - minute of the hour (0-59).
32775 (SEC) - seconds of the minute (0-59).

TiVo Receiver State:

32800 (SWREL) - software release (in x.x.x notation in bytes).
32801 (NTWRK) - object ID of currently tuned network.
32802 (PRGRM) - object ID of currently tuned program.
32803 (PSTATE) - current state of output pipeline:
0 - normal playback
1 - paused
2 - slo-mo
10 - rewind speed 1
11 - rewind speed 2
...
20 - ff speed 1
21 - ff speed 2
...

Tag Execution State:

32900 (IND) - indicator number to display or take down.

32901 (PDURING) - state of the pipeline while tag is executing.
32902 (ALTP) - alternate program object ID to push on play stack.
32903 (SELOBJ) - program object ID to record if indicator selected.

33000 (MENU1) - string object number for menu item 1.
33001 (MENU2) - string object number for menu item 2.
...
33009 (MENU10) - string object number for menu item 9.

33100 (PICT1) - picture object number for menu item 1.
33101 (PICT2) - picture object number for menu item 2.
...
33109 (PICT10) - picture object number for menu item 10.

33200 (MSELOBJ1) - program object ID to record if menu item selected.
33201 (MSELOBJ2) - program object ID to record if menu item selected.
...
33209 (MSELOBJ10) - program object ID to record if menu item selected.

Tags

- Push Alternate Program
- Pop Alternate Program (auto-pop at end of program)
- Raise Indicator
- Lower Indicator
- Menu

Tag Execution Policy

Execution policy is determined by the TSM. Some suggestions are:

- Menus

Menus are laid out as per standard TiVo menu guidelines. In general, menus appear over live video. Selection of an item typically invokes the record dialog. It may be best to pause the pipeline during the menu operation.

- Indicators

With respect to Figs. 17 and 22, indicators 1702 are lined up at the bottom of the display as small icons. During the normal viewing state, the up arrow and down arrow keys 2203 on the remote control 2201 do nothing. For indicators, up arrow 2203 circles through the indicators to the left, down arrow to the right. The selected indicator has a small square drawn around it. Pushing select 2204 initiates the action. New indicators are by default selected; if an indicator is removed, the previously selected indicator is highlighted, if any.

- Alternate Programs

Alternate programs should appear as part of the video stream, and have full ff/rew controls. The skip to live button 2202 pops the alternate program stack to empty first.

One skilled in the art will readily appreciate that although the closed caption stream is specifically mentioned above, other transport methods can be used such as the EDS fields, VBI, MPEG2 private data channel, etc.

Although the invention is described herein with reference to the preferred embodiment, one skilled in the art will readily appreciate that other applications may be substituted for those set forth herein without departing from the spirit and scope of the present invention. Accordingly, the invention should only be limited by the Claims included below.

CLAIMS

1. A process for frame specific tagging of television audio and video broadcast streams with tag translation at a receiver, comprising the steps of:

providing a storage device on said receiver;

inserting tags into said broadcast stream;

tuning said receiver to said broadcast stream;

receiving said broadcast stream at said receiver;

storing said broadcast stream on said storage device;

detecting said tags in said broadcast stream;

processing said tags;

displaying program material in said broadcast stream from said storage device to a viewer;

wherein said processing step performs the appropriate actions in response to said tags; and

wherein said tags include command and control information.

2. The process of claim 1, wherein tags indicate the start and end points of a program segment.

3. The process of claim 2, wherein said displaying step skips over said program segment in response to the viewer pressing a button on a remote input device.

4. The process of claim 2, wherein said displaying step automatically skips said program segment.

5. The process of claim 1, wherein said processing step displays a menu to the viewer based on information included in a tag.

6. The process of claim 1, wherein said processing step records the current program in said broadcast stream on said storage device based on information included in a tag.

7. The process of claim 1, said processing step further comprising the steps of:

displaying multiple icons to the viewer;

accepting viewer input information;
allowing the viewer to scroll through said multiple icons;
selecting a particular icon based on the viewer's input; and
performing an action associated with the selected icon.

8. The process of claim 1, further comprising the steps of:
wherein said processing step displays an icon to the viewer based on information included in a tag;
accepting viewer input information;
interacting with the viewer based on the tag information;
wherein said displaying step saves the exit point in the program material; and
wherein the viewer is returned to said exit point upon completion of any interaction.

9. The process of claim 8, further comprising the steps of:
presenting a plurality of menus to the viewer for generating a lead; and
forwarding the viewer's contact information to a third party upon viewer approval.

10. The process of claim 8, further comprising the steps of:
presenting a plurality of menus to the viewer for generating a sale of an advertised product or service; and
forwarding the viewer's purchase information to the proper merchant.

11. The process of claim 8, further comprising the step of:
presenting a set of program recording options to the viewer; and
scheduling the viewer's recording preferences.

12. The process of claim 8, further comprising the step of:
presenting the content of a Web site's Web page to the viewer in response to the viewer's input; and
wherein the viewer is allowed to interact with said Web site.

13. The process of claim 1, wherein said tags allow a system administrator to remotely configure said receiver.

14. The process of claim 1, further comprising the steps of:

marking indexes in said program material based on tag information; and
jumping to an index selected by the viewer.

15. A process for scheduling the recording of a television program via an advertisement in a television broadcast stream, comprising the steps of:
receiving said television broadcast stream;
playing a promotional advertisement in said television broadcast stream for a future showing of a program;
displaying an icon notifying the viewer that said program is available to record;
accepting the viewer's single key press from a remote input device; and
scheduling the recording of said program.

16. The process of claim 15, wherein said icon is displayed based on a tag inserted into said television broadcast stream.

17. The process of claim 15, further comprising the step of:
providing a storage device; and
wherein said program is stored on said storage device when the scheduled time arrives.

18. A process for the automatic replacement of program segments in a multimedia television broadcast stream at a receiver, comprising the steps of:
receiving said multimedia television broadcast stream;
detecting the start and end points of an old program segment in said broadcast stream;
providing a plurality of new program segments; and
substituting said old program segment with a new program segment during playback of said broadcast stream to a viewer.

19. The process of claim 18, wherein said detecting step searches for tags inserted into said broadcast stream denoting the start and end points of program segments.

20. The process of claim 19, wherein said tags are located in the closed caption area of said broadcast stream.

21. The process of claim 18, further comprising the step of:

providing a storage device on said receiver; and
wherein said new program segments are stored on said storage device.

22. The process of claim 21, further comprising the steps of:
receiving new program segments via said broadcast stream; and
storing said new program segments on said storage device.

23. The process of claim 18, wherein said new program segments are stored at a remotely accessible location.

24. The process of claim 18, wherein said new program segment to be played back is selected based on criteria such as: locale, the time of day, program material, the viewer's viewing habits, the viewer's program preferences, or the viewer's personal information.

25. The process of claim 24, wherein said criteria may result in the old program segment not being substituted.

26. The process of claim 24, wherein said new program segments have program objects describing their features which are used to select the best matching new program segment.

27. The process of claim 18, wherein a rotation mechanism is used when selecting said new program segments to avoid ad burnout.

28. An apparatus for frame specific tagging of television audio and video broadcast streams with tag translation at a receiver, comprising:
a storage device on said receiver;
a module for inserting tags into said broadcast stream;
a module for tuning said receiver to said broadcast stream;
a module for receiving said broadcast stream at said receiver;
a module for storing said broadcast stream on said storage device;
a module for detecting said tags in said broadcast stream;
a module for processing said tags;
a module for displaying program material in said broadcast stream from said storage device to a viewer;
wherein said processing module performs the appropriate actions in response to said tags; and

wherein said tags include command and control information.

29. The apparatus of claim 28, wherein tags indicate the start and end points of a program segment.

30. The apparatus of claim 29, wherein said displaying module skips over said program segment in response to the viewer pressing a button on a remote input device.

31. The apparatus of claim 29, wherein said displaying module automatically skips said program segment.

32. The apparatus of claim 28, wherein said processing module displays a menu to the viewer based on information included in a tag.

33. The apparatus of claim 28, wherein said processing module records the current program in said broadcast stream on said storage device based on information included in a tag.

34. The apparatus of claim 28, said processing module further comprising:
a module for displaying multiple icons to the viewer;
a module for accepting viewer input information;
a module for allowing the viewer to scroll through said multiple icons;
a module for selecting a particular icon based on the viewer's input; and
a module for performing an action associated with the selected icon.

35. The apparatus of claim 28, further comprising:
wherein said processing module displays an icon to the viewer based on information included in a tag;
a module for accepting viewer input information;
a module for interacting with the viewer based on the tag information;
wherein said displaying module saves the exit point in the program material; and
wherein the viewer is returned to said exit point upon completion of any interaction.

36. The apparatus of claim 35, further comprising:
a module for presenting a plurality of menus to the viewer for generating a

lead; and

a module for forwarding the viewer's contact information to a third party upon viewer approval.

37. The apparatus of claim 35, further comprising:

a module for presenting a plurality of menus to the viewer for generating a sale of an advertised product or service; and

a module for forwarding the viewer's purchase information to the proper merchant.

38. The apparatus of claim 35, further comprising:

a module for presenting a set of program recording options to the viewer; and

a module for scheduling the viewer's recording preferences.

39. The apparatus of claim 35, further comprising:

a module for presenting the content of a Web site's Web page to the viewer in response to the viewer's input; and

wherein the viewer is allowed to interact with said Web site.

40. The apparatus of claim 28, wherein said tags allow a system administrator to remotely configure said receiver.

41. The apparatus of claim 28, further comprising:

a module for marking indexes in said program material based on tag information; and

a module for jumping to an index selected by the viewer.

42. An apparatus for scheduling the recording of a television program via an advertisement in a television broadcast stream, comprising:

a module for receiving said television broadcast stream;

a module for playing a promotional advertisement in said television broadcast stream for a future showing of a program;

a module for displaying an icon notifying the viewer that said program is available to record;

a module for accepting the viewer's single key press from a remote input device; and

a module for scheduling the recording of said program.

43. The apparatus of claim 42, wherein said icon is displayed based on a tag inserted into said television broadcast stream.

44. The apparatus of claim 42, further comprising:

a storage device; and

wherein said program is stored on said storage device when the scheduled time arrives.

45. A apparatus for the automatic replacement of program segments in a multimedia television broadcast stream at a receiver, comprising:

a module for receiving said multimedia television broadcast stream;

a module for detecting the start and end points of an old program segment in said broadcast stream;

a module for providing a plurality of new program segments; and

a module for substituting said old program segment with a new program segment during playback of said broadcast stream to a viewer.

46. The apparatus of claim 45, wherein said detecting module searches for tags inserted into said broadcast stream denoting the start and end points of program segments.

47. The apparatus of claim 46, wherein said tags are located in the closed caption area of said broadcast stream.

48. The apparatus of claim 45, further comprising:

a storage device on said receiver; and

wherein said new program segments are stored on said storage device.

49. The apparatus of claim 48, further comprising:

a module for receiving new program segments via said broadcast stream; and

a module for storing said new program segments on said storage device.

50. The apparatus of claim 45, wherein said new program segments are stored at a remotely accessible location.

51. The apparatus of claim 45, wherein said new program segment to be

played back is selected based on criteria such as: locale, the time of day, program material, the viewer's viewing habits, the viewer's program preferences, or the viewer's personal information.

52. The apparatus of claim 51, wherein said criteria may result in the old program segment not being substituted.

53. The apparatus of claim 51, wherein said new program segments have program objects describing their features which are used to select the best matching new program segment.

54. The apparatus of claim 45, wherein a rotation mechanism is used when selecting said new program segments to avoid ad burnout.

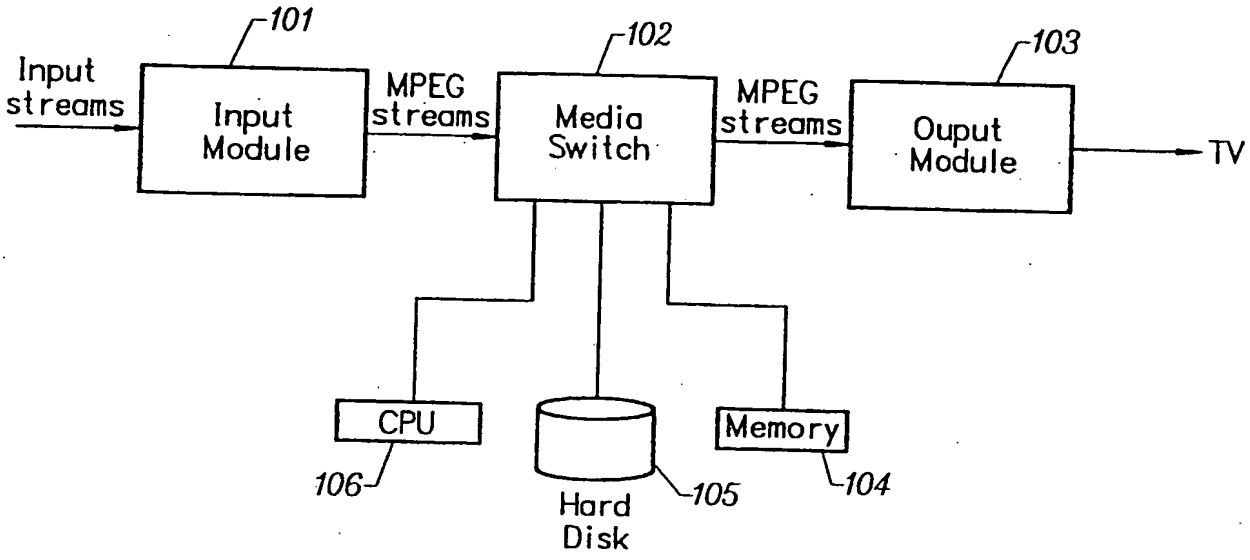


FIG. 1

INSDOCID: <WO_____0122729A1_I_>

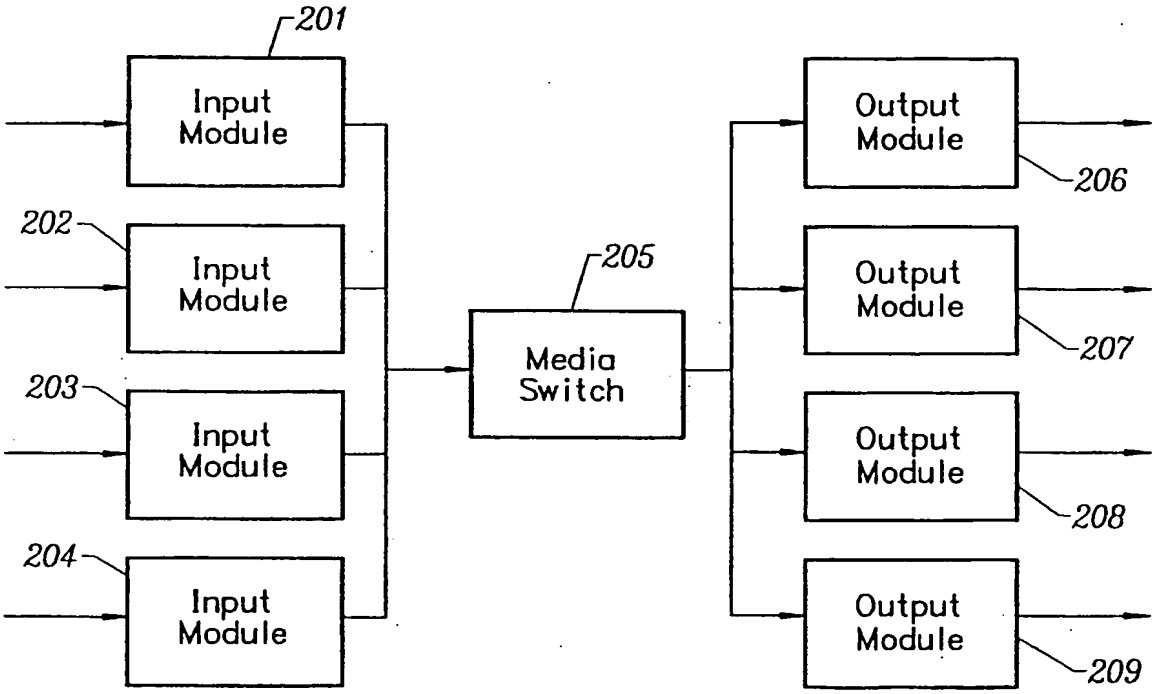

201
Input Module
202
Input Module
203
Input Module
204
Input Module
205
Media Switch
Output Module
206
Output Module
207
Output Module
208
Output Module
209

FIG. 2

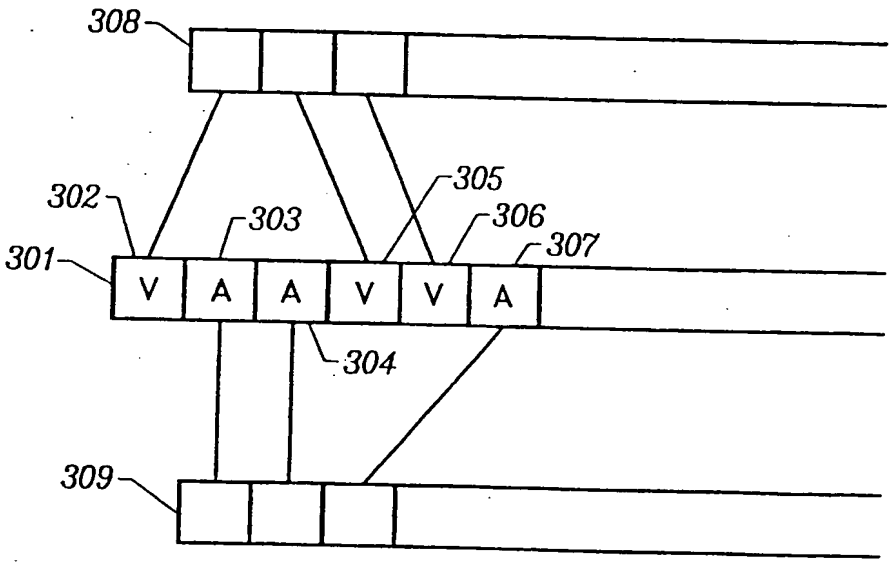


FIG. 3

BNSDOCID: <WO_____0122729A1_I_>

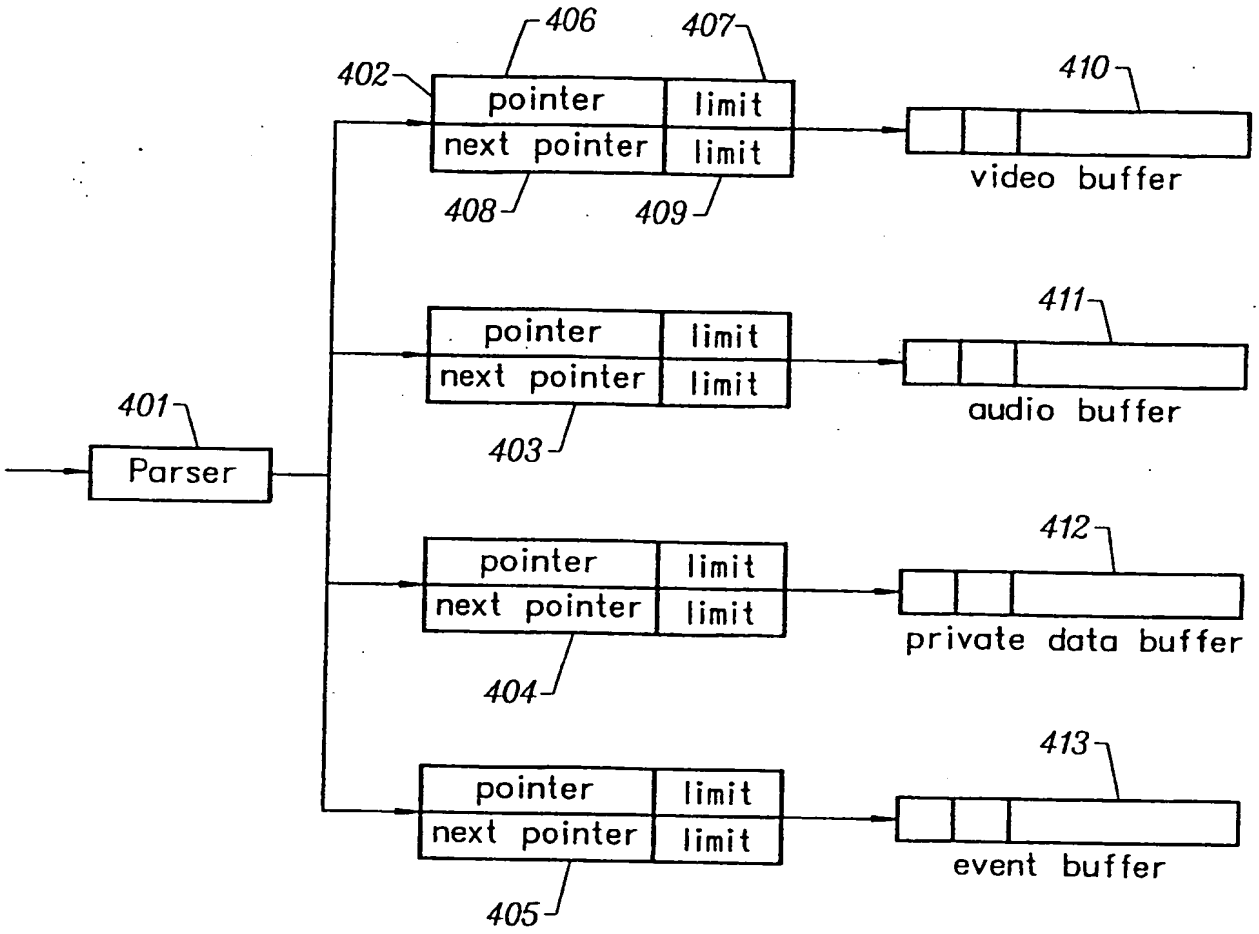


FIG. 4

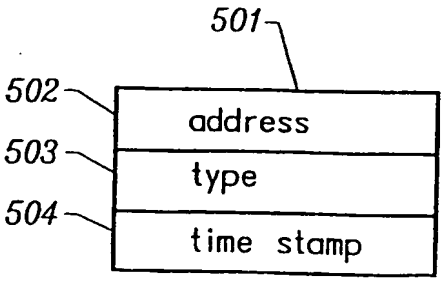


FIG. 5

BNSDOCID: <WO_____0122729A1_I_>

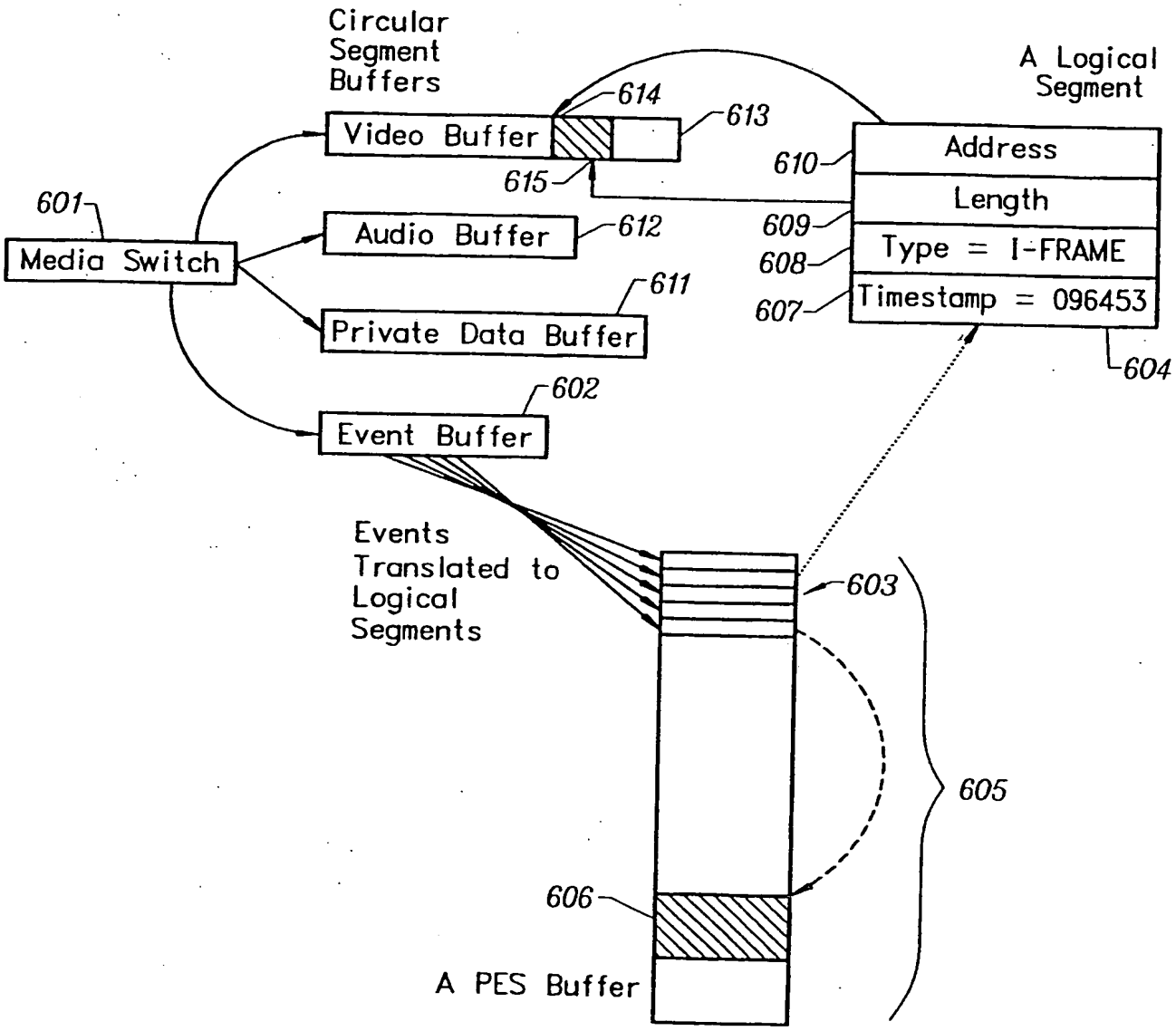

Circular
Segment
Buffers
A Logical
Segment
614
613
Video Buffer
615
610
Address
Length
609
Type = I-FRAME
608
Timestamp = 096453
607
604
601
Media Switch
Audio Buffer
612
611
Private Data Buffer
602
Event Buffer
Events
Translated to
Logical
Segments
603
605
606
A PES Buffer

FIG. 6

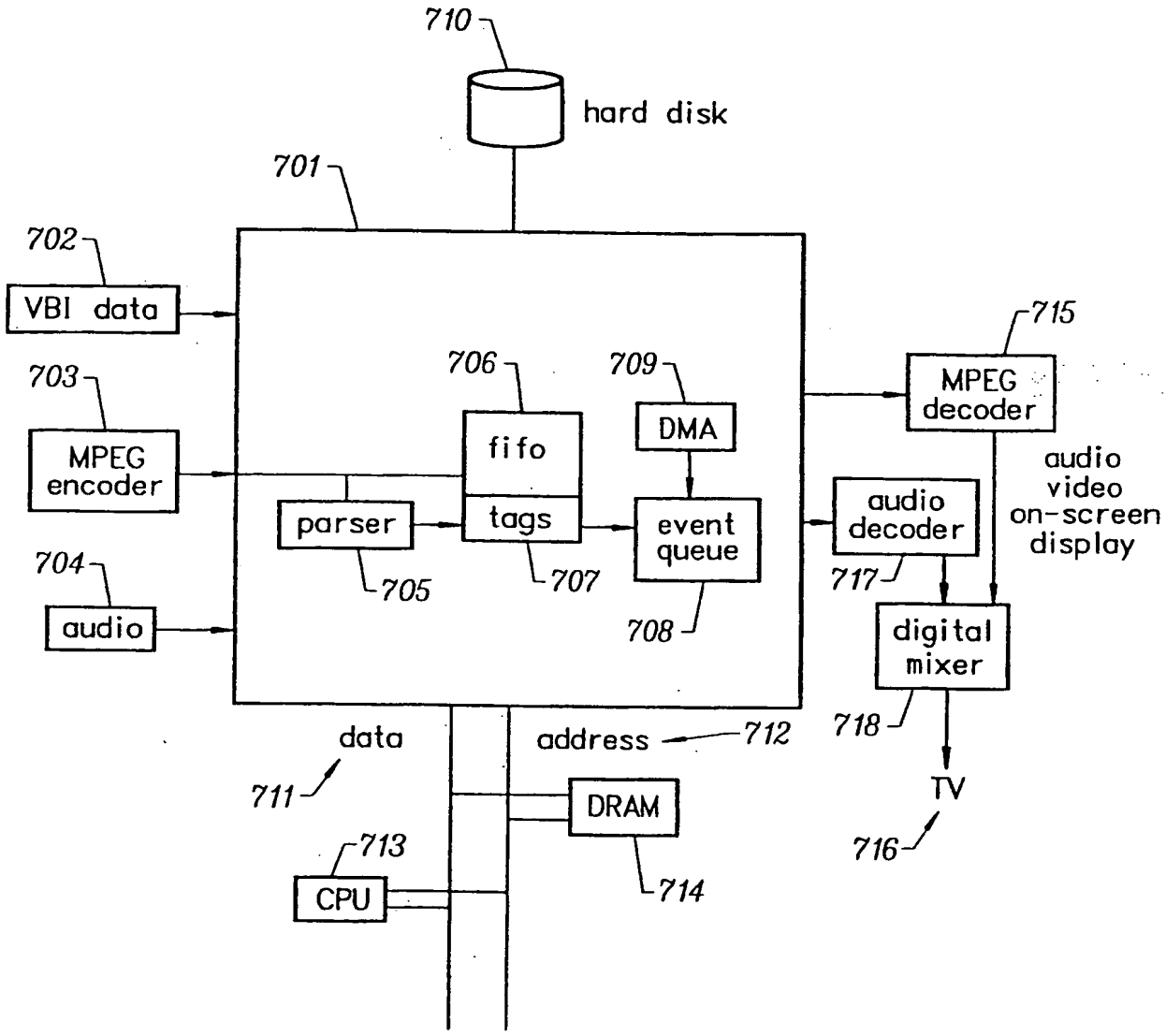


FIG. 7

BNSDOCID: <WO_____0122729A1_I_>

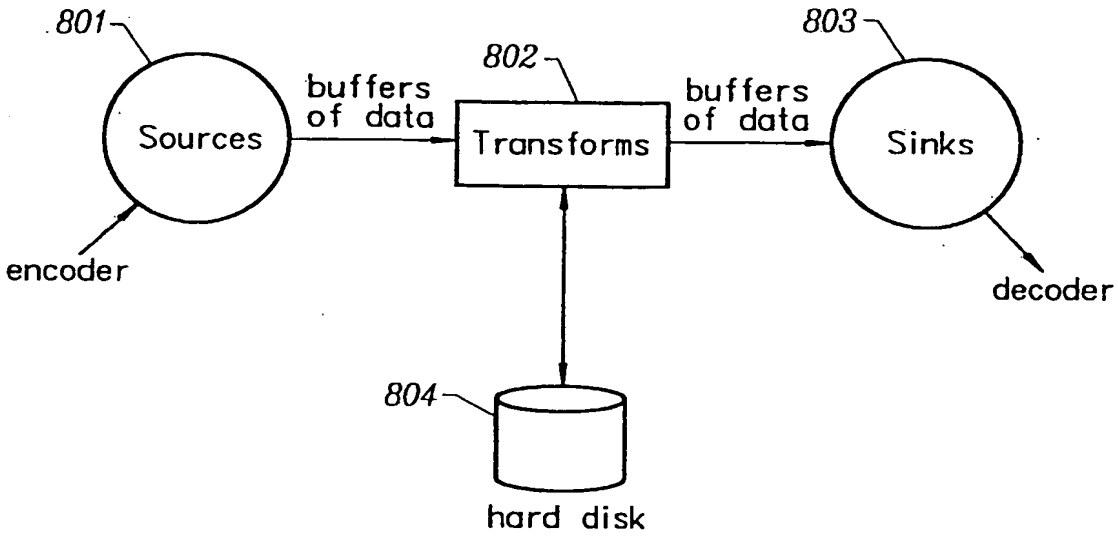


FIG. 8

BNSDOCID: <WO_____0122729A1_I_>

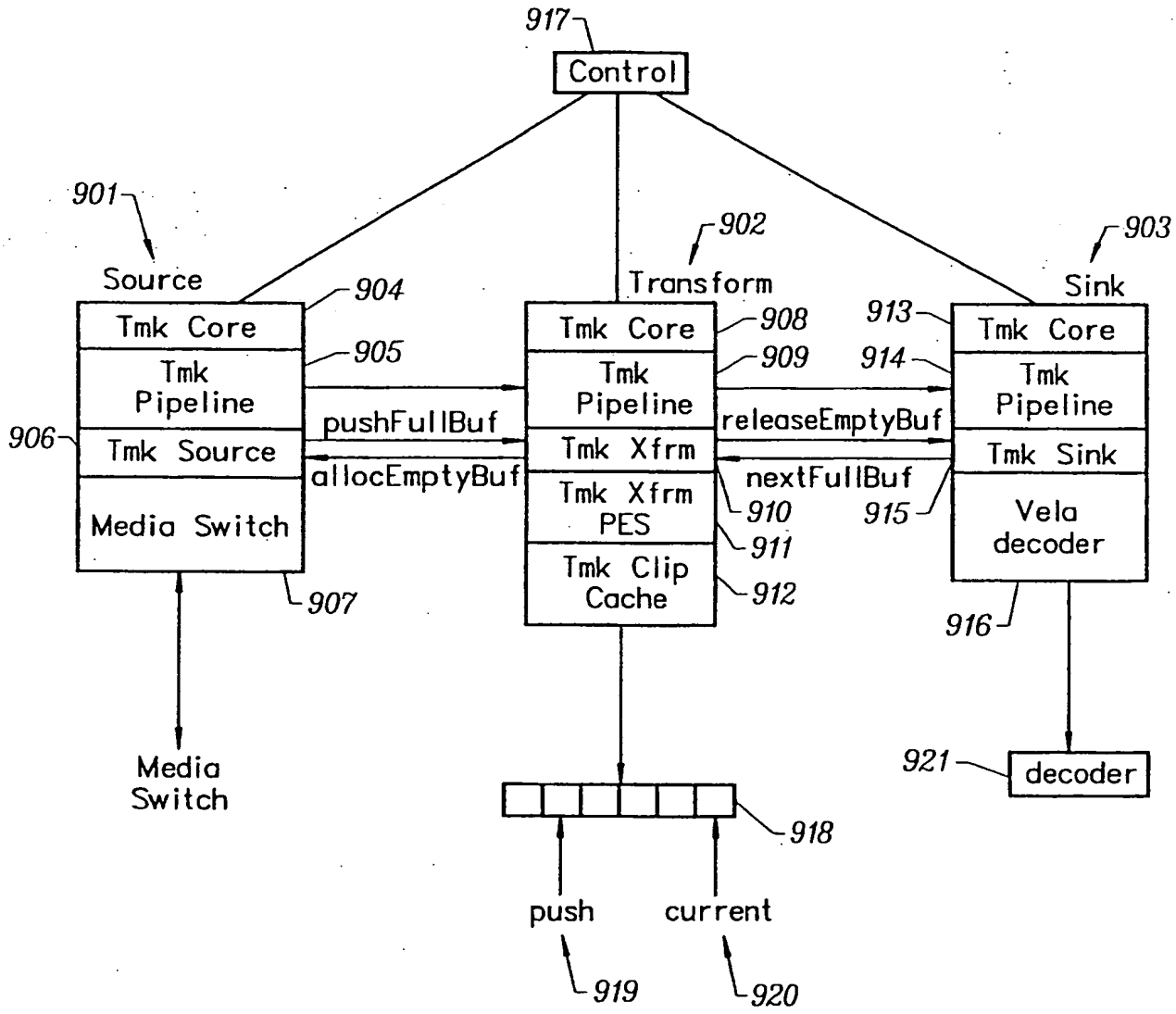


FIG. 9

BNSDOCID: <WO_____0122729A1_I_>

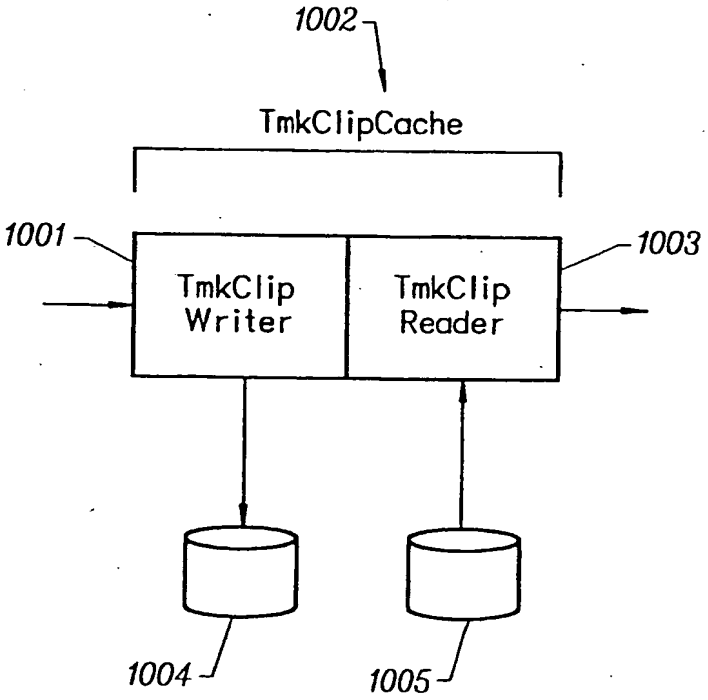

1002
TmkClipCache
1001
TmkClip Writer
TmkClip Reader
1003
1004
1005

FIG. 10

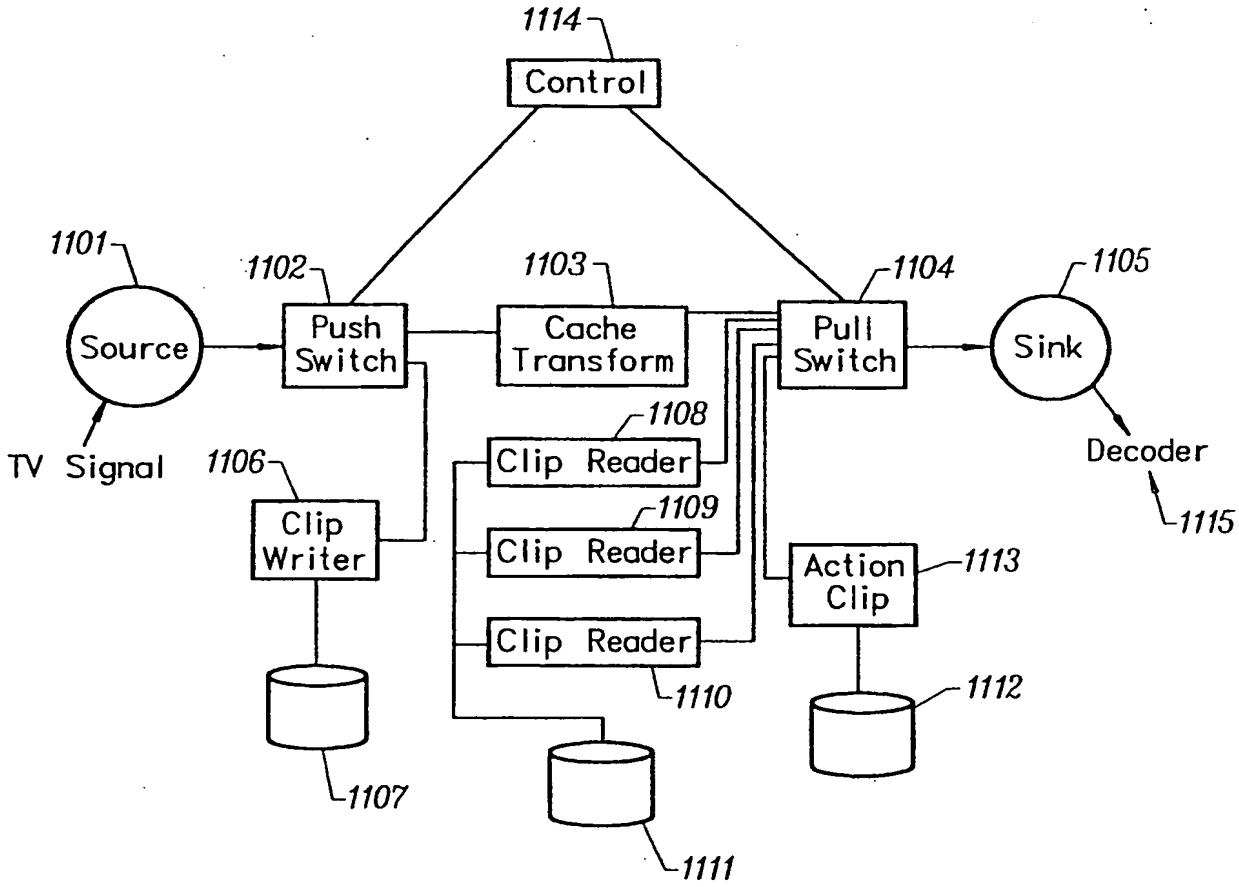


FIG. 11

BNSDOCID: <WO_____0122729A1_I_>

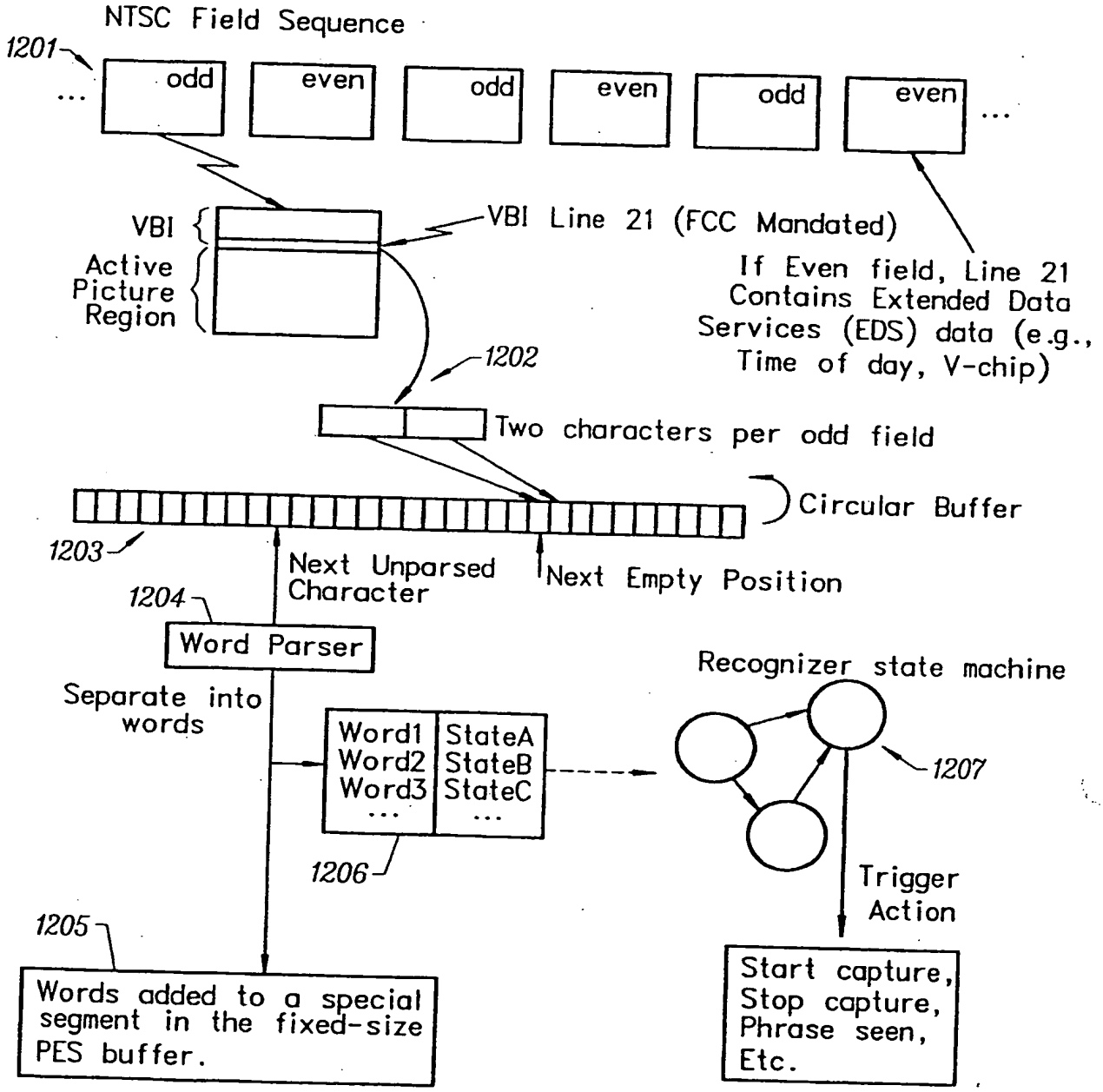


FIG. 12

3NSDOCID: <WO_____0122729A1_I_>

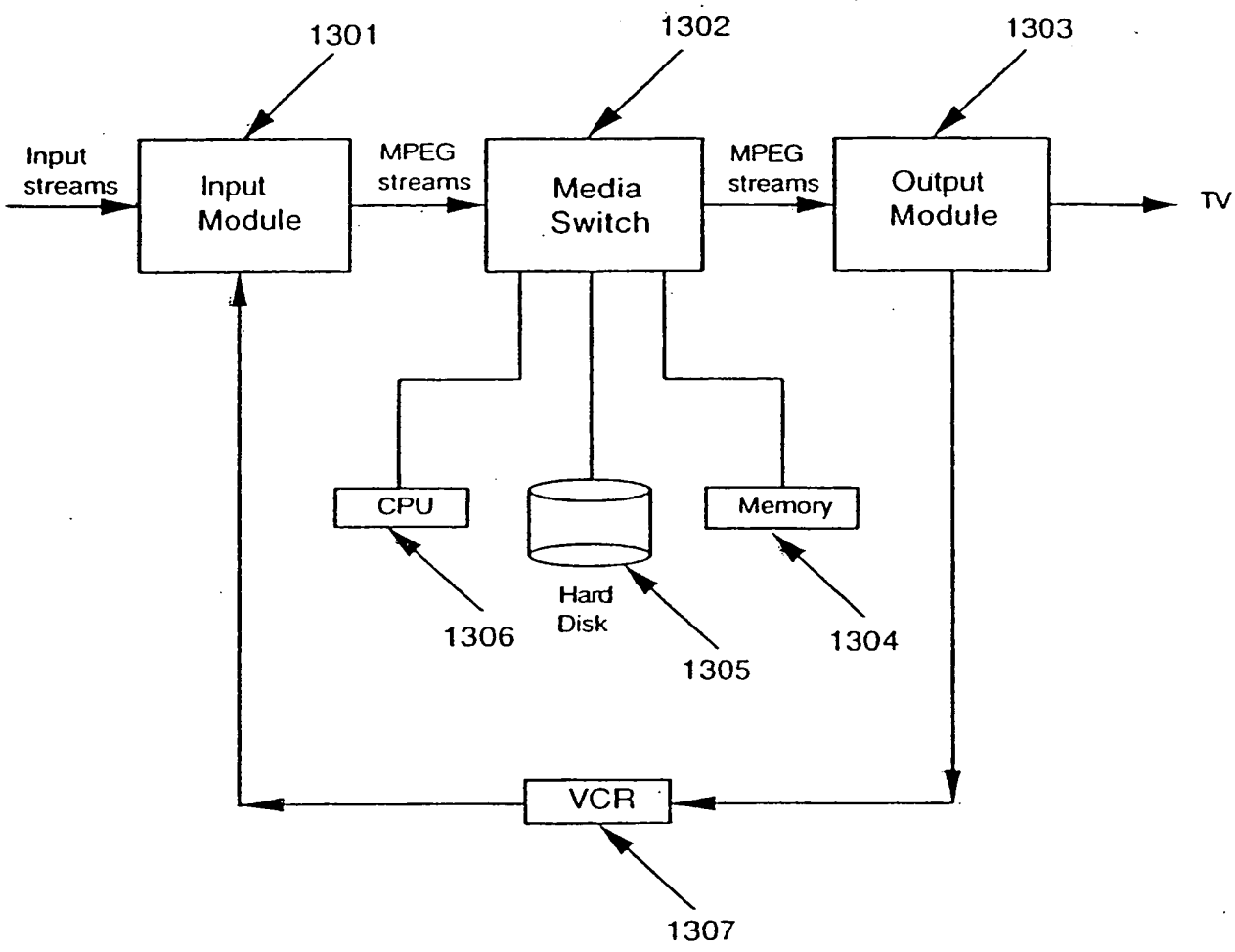


Fig. 13

BNSDOCID: <WO_____0122729A1_I_>

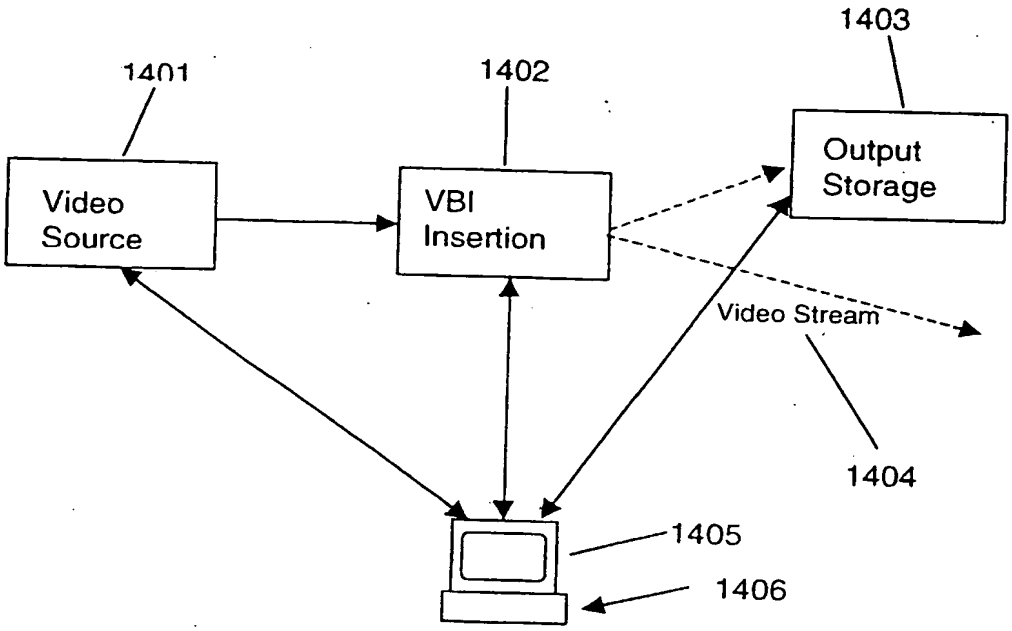

1403
1401
1402
Output Storage
Video Source
VBI Insertion
Video Stream
1404
1405
1406

Fig. 14

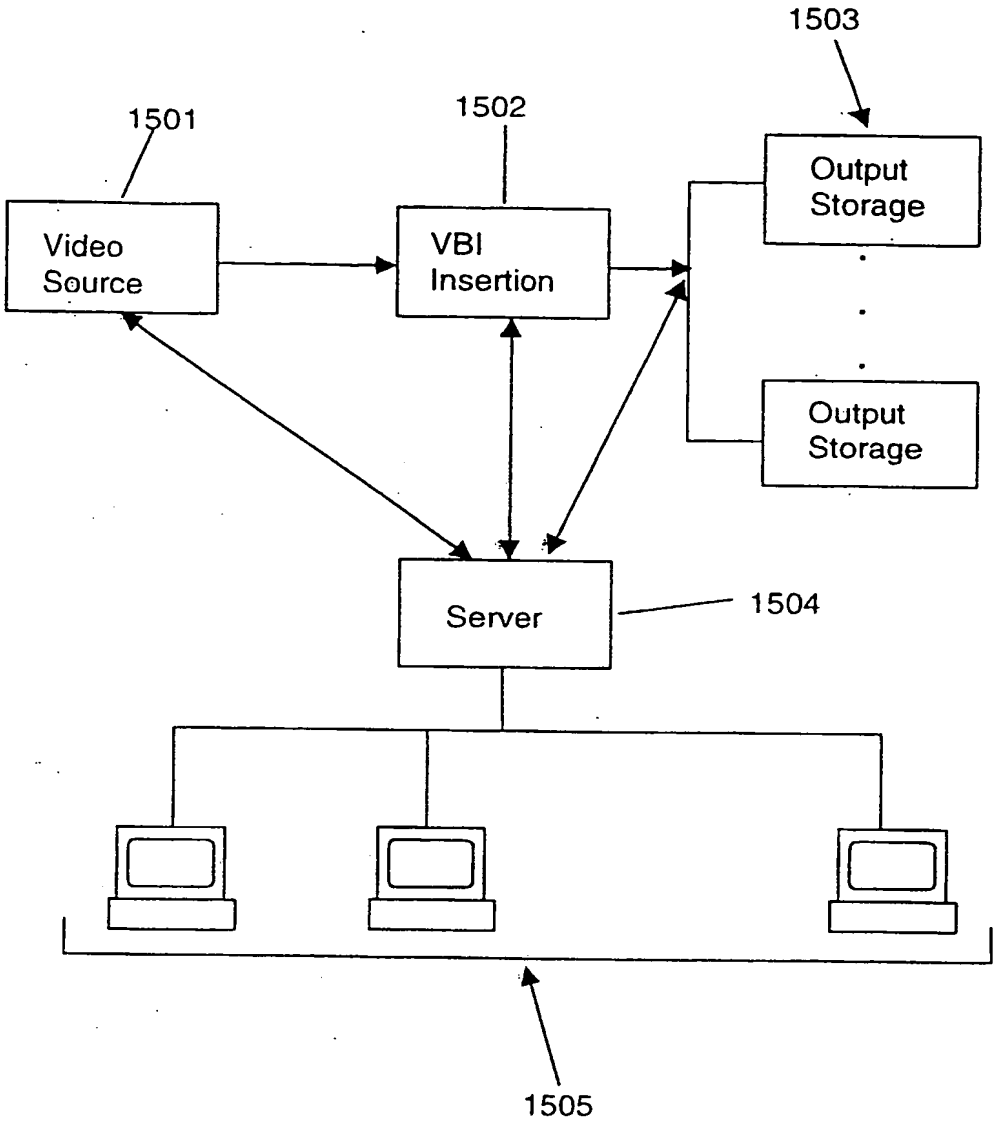

1503
1501
1502
Output Storage
Video Source
VBI Insertion
Output Storage
Server
1504
1505

Fig. 15

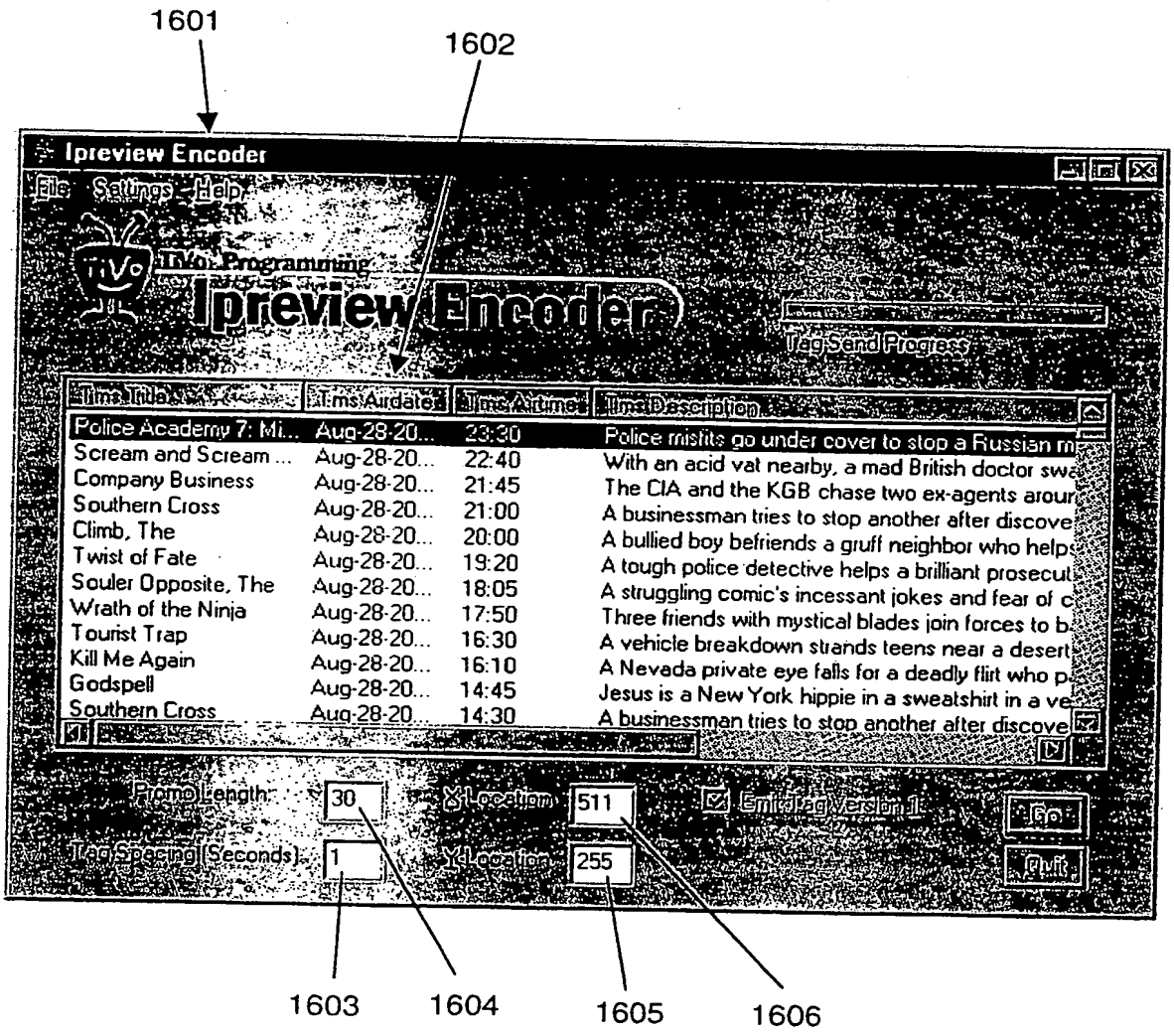

1601
1602
Ipreview Encoder
File Settings Help
TiVo Programming
Ipreview Encoder
Tag Send Progress
Police Academy 7: Mi... Aug-28-20... 23:30 Police misfits go under cover to stop a Russian m
Scream and Scream ... Aug-28-20... 22:40 With an acid vat nearby, a mad British doctor swa
Company Business Aug-28-20... 21:45 The CIA and the KGB chase two ex-agents arour
Southern Cross Aug-28-20... 21:00 A businessman tries to stop another after discove
Climb, The Aug-28-20... 20:00 A bullied boy befriends a gruff neighbor who help
Twist of Fate Aug-28-20... 19:20 A tough police detective helps a brilliant prosecut
Souler Opposite, The Aug-28-20... 18:05 A struggling comic's incessant jokes and fear of c
Wrath of the Ninja Aug-28-20... 17:50 Three friends with mystical blades join forces to b
Tourist Trap Aug-28-20... 16:30 A vehicle breakdown strands teens near a desert
Kill Me Again Aug-28-20... 16:10 A Nevada private eye falls for a deadly flirt who p
Godspell Aug-28-20... 14:45 Jesus is a New York hippie in a sweatshirt in a ve
Southern Cross Aug-28-20... 14:30 A businessman tries to stop another after discove
Promo Length: 30
X Location: 511
Emit Tag Version 1
Go
Tag Spacing (Seconds): 1
Y Location: 255
Quit
1603
1604
1605
1606

Fig. 16

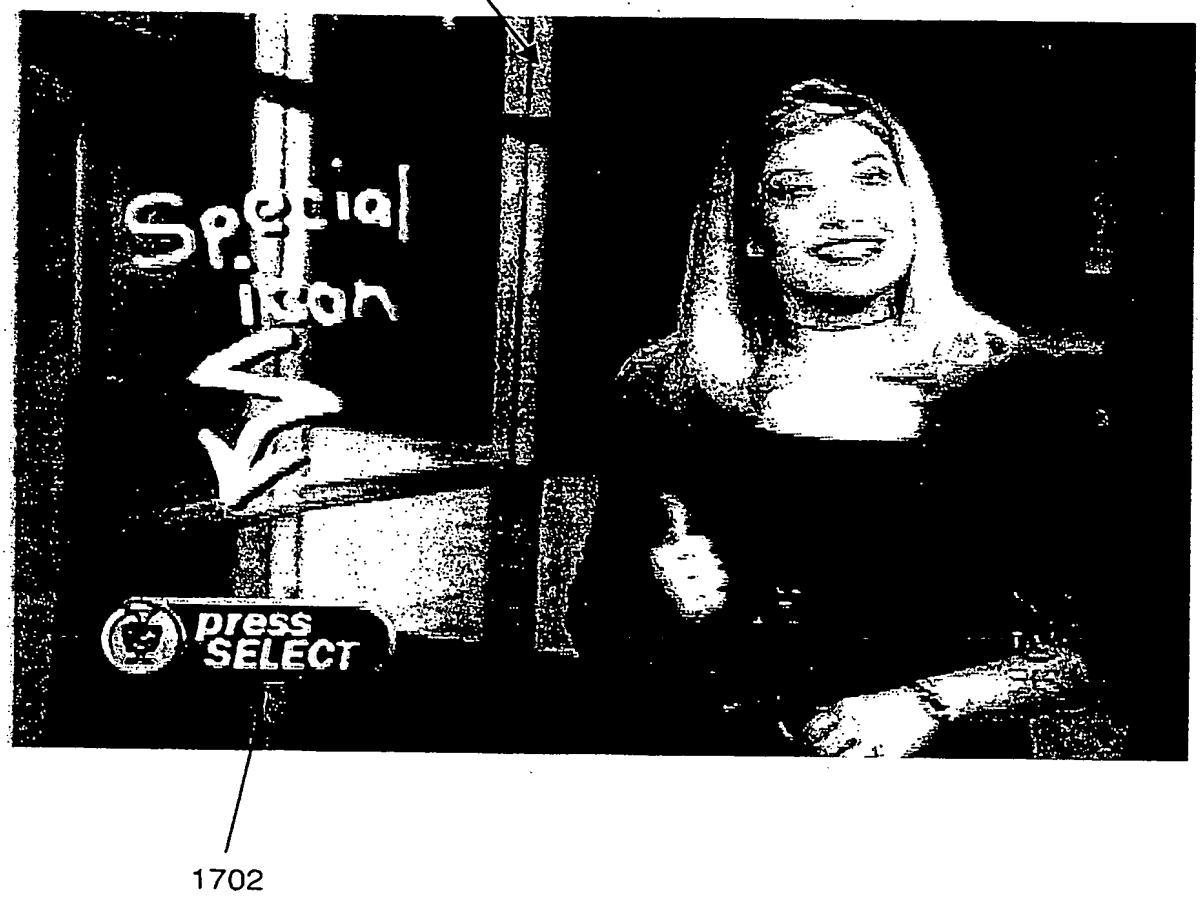


Fig. 17

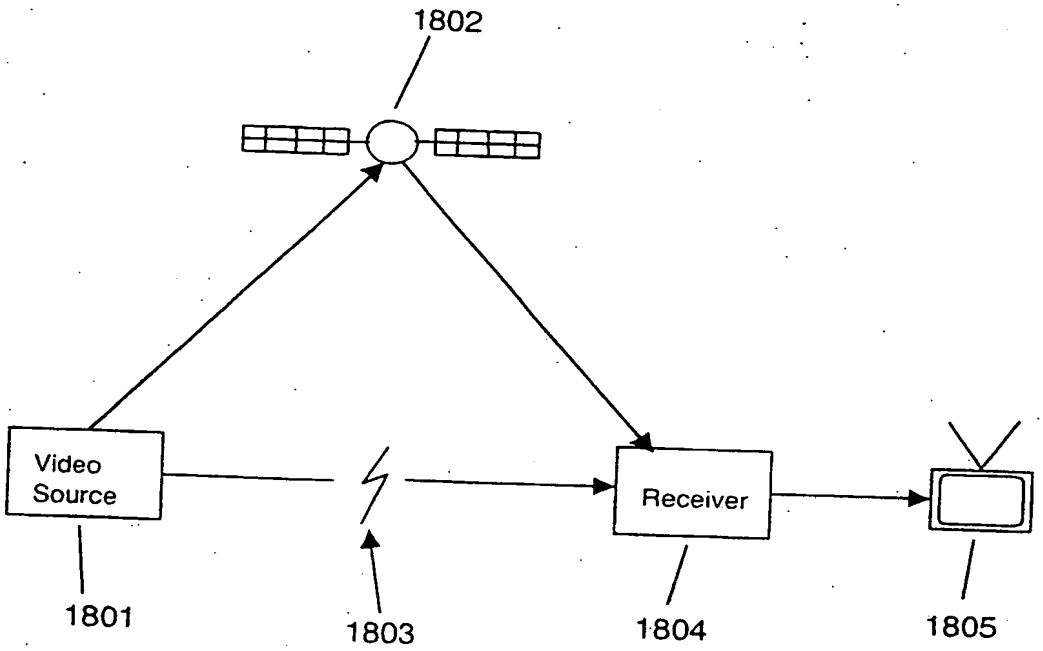


Fig. 18

NSDOCID: <WO_____0122729A1_I_>

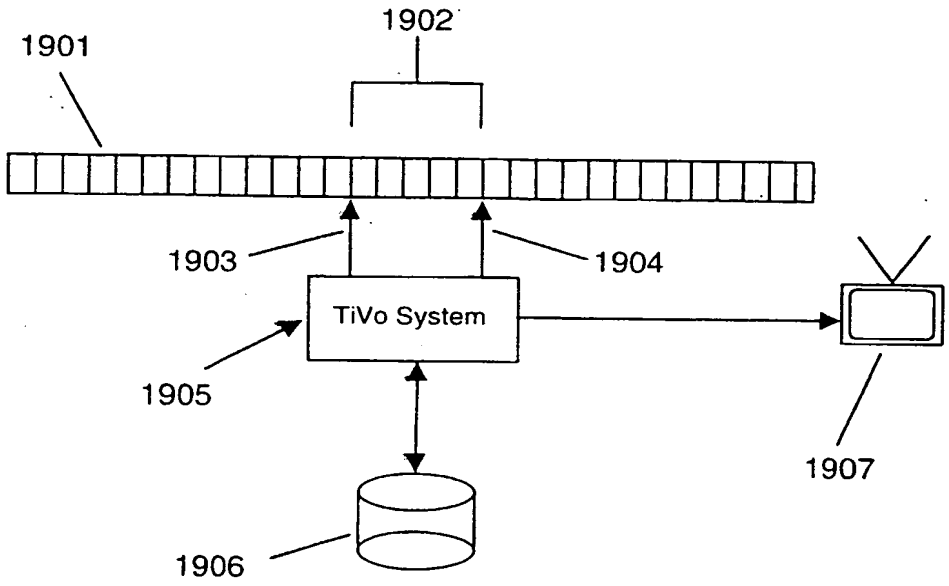


Fig. 19

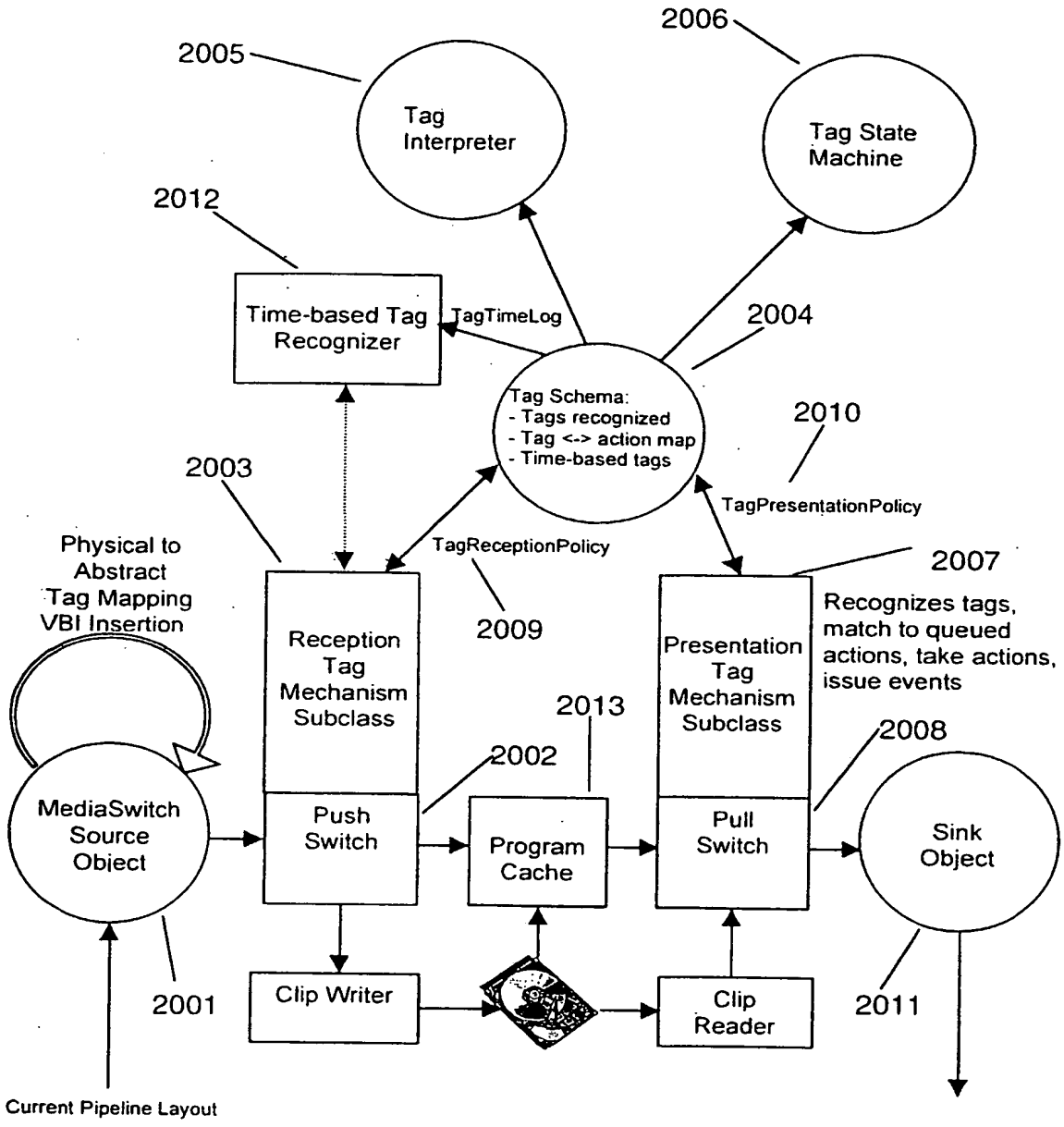

2005
Tag Interpreter
2006
Tag State Machine
2012
Time-based Tag Recognizer
TagTimeLog
2004
Tag Schema:
- Tags recognized
- Tag <-> action map
- Time-based tags
2010
TagPresentationPolicy
2003
TagReceptionPolicy
2009
2007
Recognizes tags, match to queued actions, take actions, issue events
Physical to Abstract Tag Mapping VBI Insertion
Reception Tag Mechanism Subclass
Presentation Tag Mechanism Subclass
2013
2008
2002
MediaSwitch Source Object
Push Switch
Program Cache
Pull Switch
Sink Object
2001
Clip Writer
Clip Reader
2011
Current Pipeline Layout

Fig. 20

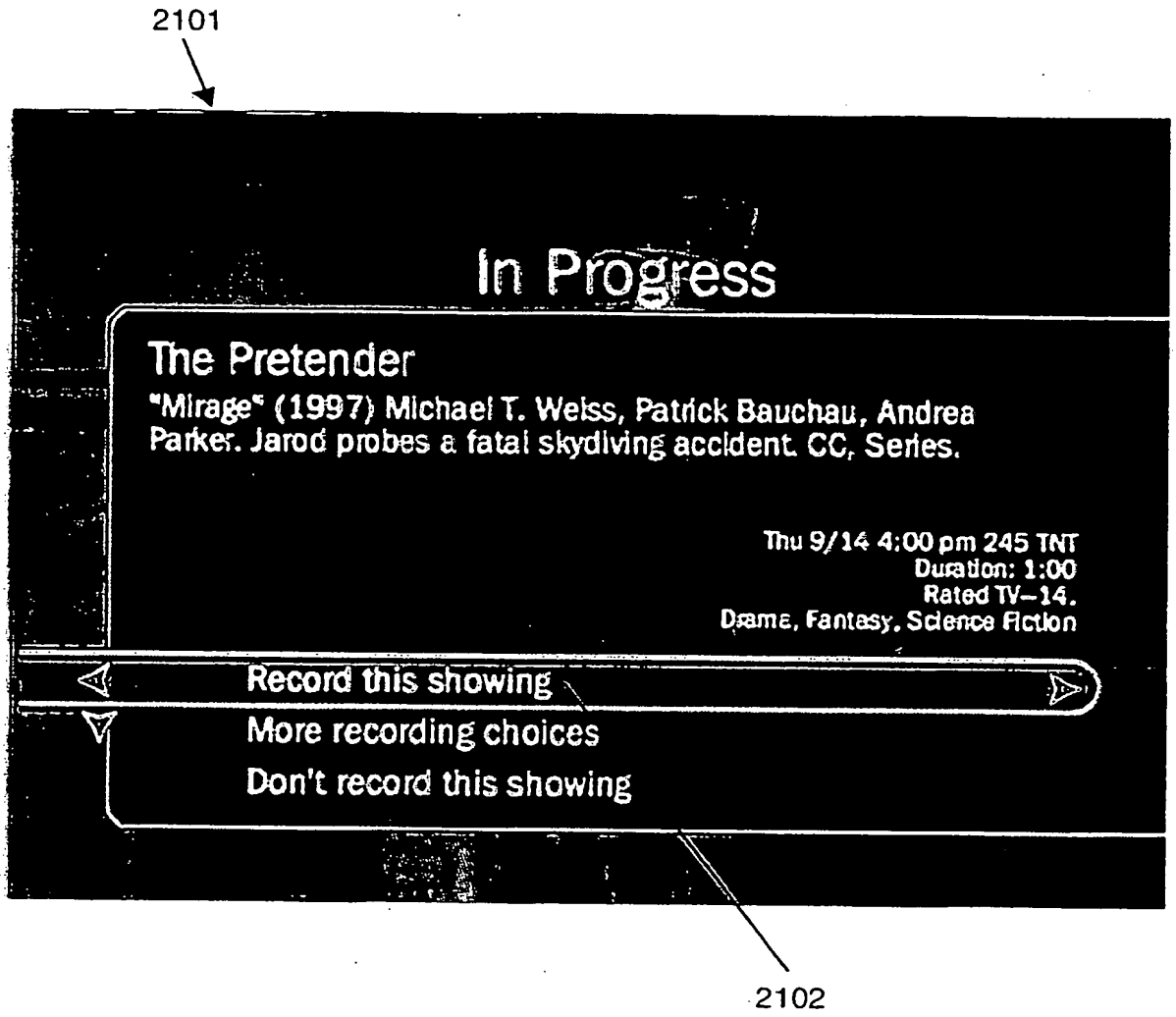

2101
In Progress
The Pretender
"Mirage" (1997) Michael T. Weiss, Patrick Bauchau, Andrea Parker. Jarod probes a fatal skydiving accident. CC, Series.
Thu 9/14 4:00 pm 245 TNT
Duration: 1:00
Rated TV-14.
Drama, Fantasy, Science Fiction
Record this showing
More recording choices
Don't record this showing
2102

Fig. 21

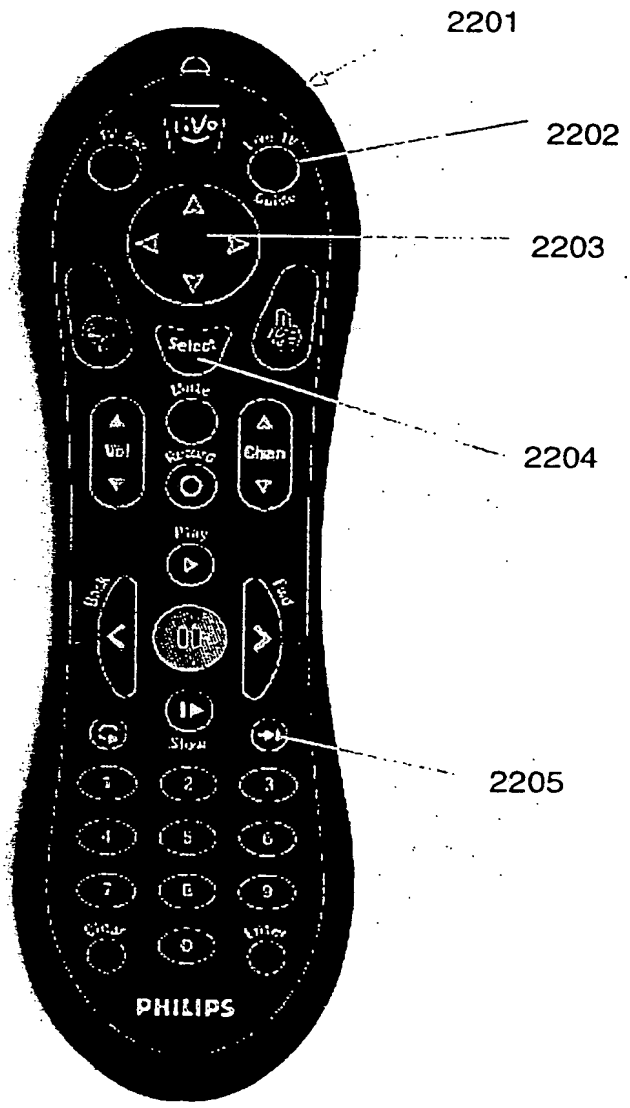


Fig. 22

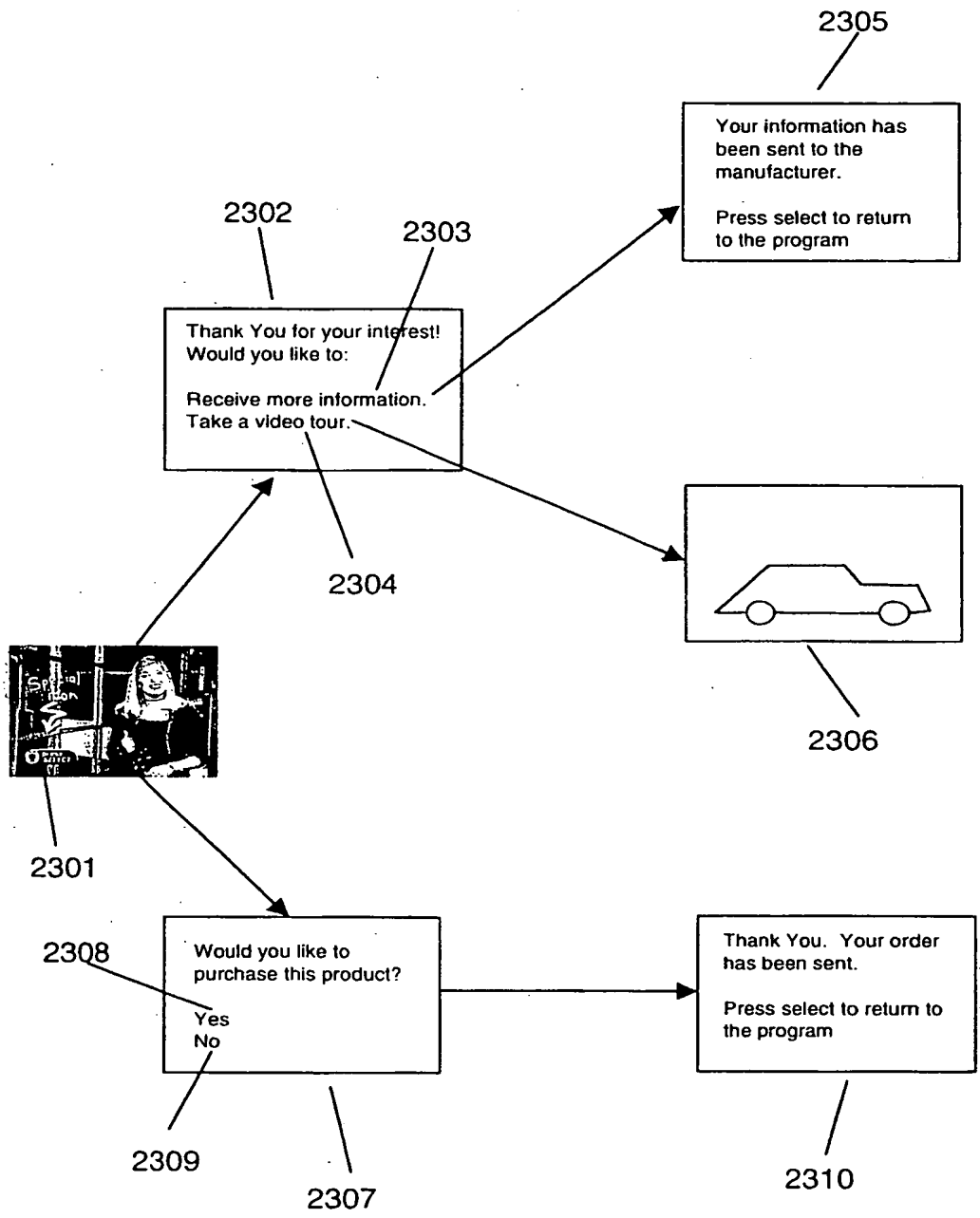

2305
Your information has been sent to the manufacturer.
Press select to return to the program
2302
2303
Thank You for your interest! Would you like to:
Receive more information.
Take a video tour.
2304
2306
2301
2308
Would you like to purchase this product?
Yes
No
Thank You. Your order has been sent.
Press select to return to the program
2309
2307
2310

Fig. 23

INTERNATIONAL SEARCH REPORT

Inte nal Application No
PCT/US 00/25847

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 H04N5/92

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 H04N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 96 08921 A (ARTHUR D. LITTLE INTERPRISES, INC) 21 March 1996 (1996-03-21) page 9, line 15 -page 27, line 20; figures 1-10	1-4, 28-31
A	DE 44 34 034 A (DEUTSCHE THOMSON-BRANDT GMBH) 28 March 1996 (1996-03-28) the whole document	1-4,6, 28-31,33
A	US 5 805 763 A (LAWLER ET AL.) 8 September 1998 (1998-09-08) the whole document	1,5-8, 10,11, 15,17, 28, 32-35, 37,38, 42,44

-/--

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
9 January 2001	15/01/2001

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Verleye, J

1

Form PCT/ISA/210 (second sheet) (July 1992)

BNSDOCID: <WO_____0122729A1_I_>

INTERNATIONAL SEARCH REPORT

Int al Application No
PCT/US 00/25847

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 5 233 423 A (JERNIGAN ET AL.) 3 August 1993 (1993-08-03) the whole document	18,45

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

BNSDOCID: <WO_____0122729A1_I_>

INTERNATIONAL SEARCH REPORT

formation on patent family members

Int· nal Application No
PCT/US 00/25847

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9608921 A	21-03-1996	US 5696866 A	09-12-1997
		AU 694568 B	23-07-1998
		AU 3680295 A	29-03-1996
		BR 9508912 A	02-06-1998
		CA 2199485 A	21-03-1996
		EP 0781486 A	02-07-1997
		JP 10507884 T	28-07-1998
		US 5987210 A	16-11-1999
		US 5999688 A	07-12-1999
DE 4434034 A	28-03-1996	NONE	
US 5805763 A	08-09-1998	NONE	
US 5233423 A	03-08-1993	NONE	

BNSDOCID: <WO_____0122728A1_I_>